UNIVERSITÉ CHARLES DE GAULLE - LILLE III

THÈSE

en vue de l'obtention du grade de

Docteur

Discipline: Mathématiques Appliquées et Applications des Mathématiques

présentée par

Fedya TELMOUDI

sous le titre

Estimation and misspecification risks in VaR evaluation

Soutenue le : 19 décembre 2014

Composition du Jury

Directeurs de thèse	Christian FRANCQ	Université Lille 3
	Mohamed LIMAM	Université de Tunis
Rapporteurs	Mohamed-Ali EL AROUI	Université de Carthage
	Genaro SUCARRAT	BI Norwegian Business School
Examinateurs	Mohamed EL GHOURABI	Université de Tunis
	Christophe HURLIN	Université d'Orléans
	Jean-Michel ZAKOÏAN	Université Lille 3

Remerciements

Au terme de ce travail, j'adresse de chaleureux remerciements à mes directeurs de thèse Messieurs les Professeurs Christian FRANCQ et Mohamed LIMAM pour la confiance qu'ils m'ont témoignée en acceptant d'encadrer avec compétence et enthousiasme mon travail de recherche tout au long de ces trois années de thèse. Pour l' ampleur et l' efficacité de leurs encadrements, leurs qualités scientifiques, leurs patiences, leurs soutiens et surtout leurs conseils très précieux dont ils ont fait preuve au cours de l' élaboration de cette thèse. Soyez assuré de mon attachement et de ma profonde gratitude.

Je remercie vivement Monsieur le Professeur Mohamed-Ali EL AROUI et Monsieur le Professeur Genaro SUCARRAT de l'honneur qu'ils me font d'être membres de mon jury de thèse en qualité des rapporteurs de cette thèse et pour l'intérêt qu'ils ont porté à mon travail. Je suis très honorée à remercier Monsieur le Professeur Jean-Michel ZAKOÏAN, Monsieur le Professeur Christophe HURLIN et Monsieur Mohamed EL GHOURABI d'avoir accepter d'être membres de mon jury de thèse. Je remercie vivement Madame Christiane FRANCQ pour le temps consacré à la lecture très attentive de la partie en français de cette thèse ainsi que pour les commentaires m' ayant permis d' améliorer mon travail. Je remercie également les membres des laboratoires EQUIPPE et LARODEC.

Je voudrais aussi adresser un grand merci à mon père Hamed TELMOUDI, ma mère Fatma TELMOUDI, mon mari Wahib ABID pour leurs soutiens qui m'a été bien utile durant ma thèse. Merci pour avoir fait de moi ce que je suis aujourd'hui.

Je dédie ce travail à mes frères, ma soeur et mes amis qui m'ont également soutenu.

Enfin, j' exprime tous mes sentiments de reconnaissance et gratitudes à tous ceux qui ont contribué de prés ou de loin au bon déroulement de ce travail.

Résumé

Dans cette thèse, nous étudions l'estimation de la valeur à risque conditionnelle (VaR) en tenant compte du risque d'estimation et du risque de modèle. Tout d'abord, nous considérons une méthode en deux étapes pour estimer la VaR. La première étape évalue le paramètre de volatilité θ_0 en utilisant un estimateur quasi maximum de vraisemblance généralisé (gQMLE) fondé sur une densité instrumentale h. La seconde étape estime un quantile des innovations η_t à partir du quantile empirique des résidus obtenus dans la première étape. Nous donnons des conditions sous lesquelles l'estimateur en deux étapes de la VaR est convergent et asymptotiquement normal. Nous comparons également les efficacités des estimateurs obtenus pour divers choix de la densité instrumentale h. Lorsque l'innovation η_t n'est pas de densité h, la première étape donne généralement un estimateur biaisé de θ_0 et la seconde étape donne aussi un estimateur biaisé du quantile de η_t. Cependant, nous montrons que les deux erreurs se contrebalancent pour donner une estimation consistante de la VaR. Nous nous concentrons ensuite sur l'estimation de la VaR dans le cadre de modèles GARCH en utilisant le gQMLE fondé sur la classe des densités instrumentales double gamma généralisées de paramètres (b,p,d). Cette classe contient la distribution gaussienne pour $b = \frac{1}{\sqrt{2}}$, $p = 1$ et $d = 2$. Notre objectif est de comparer la performance du QMLE gaussien par rapport à celle du gQMLE. Le choix de l'estimateur optimal dépend essentiellement du paramètre d qui minimise la variance asymptotique. Nous testons si le paramètre d qui minimise la variance asymptotique est égal à 2. Lorsque le test est appliqué sur des séries réelles de rendements financiers, l'hypothèse stipulant l'optimalité du QMLE gaussien est généralement rejetée. Finalement, nous considérons les méthodes non-paramétriques d'apprentissage automatique pour estimer la VaR. Ces méthodes visent à s'affranchir du risque de modèle car elles ne reposent pas sur une forme spécifique de la volatilité. Nous utilisons la technique des machines à vecteurs de support pour la régression (SVR) basée sur la fonction de perte moindres carrés (en anglais LS). Pour améliorer la solution du modèle LS-SVR nous utilisons les modèles LS-SVR pondérés et LS-SVR de taille fixe. Des illustrations numériques mettent en évidence l'apport des modèles proposés pour estimer la VaR en tenant

compte des risques de spécification et d'estimation.

Mots clés. Estimateur efficace, Modèles d'apprentissage automatique, Modèles GARCH, Quasi maximum de vraisemblance généralisé, Risque d'estimation, Risque de mauvaise spécification, Risque de modèle, VaR conditionnelle.

Abstract

In this thesis, we study the problem of conditional Value at Risk (VaR) estimation taking into account estimation risk and model risk. First, we considered a two-step method for VaR estimation. The first step estimates the volatility parameter θ_0 using a generalized quasi maximum likelihood estimator (gQMLE) based on an instrumental density h. The second step estimates a quantile of innovations η_t from the empirical quantile of residuals obtained in the first step. We give conditions under which the two-step estimator of the VaR is consistent and asymptotically normal. We also compare the efficiencies of the estimators for various instrumental densities h. When the distribution of η_t is not the density h, the first step usually gives a biased estimator of θ_0 and the second step gives a biased estimator of the quantile of η_t. However, we show that both errors counterbalance each other to give a consistent estimate of the VaR. We then focus on the VaR estimation within the framework of GARCH models using the gQMLE based on a class of instrumental densities called double generalized gamma with parameters (b, p, d). This class contains the Gaussian distribution for $b = \frac{1}{\sqrt{2}}$, $p = 1$ and $d = 2$. Our goal is to compare the performance of the Gaussian QMLE against the gQMLE. The choice of the optimal estimator depends on the value d_0 of d that minimizes the asymptotic variance. We test if d_0 is equal 2. When the test is applied to real series of financial returns, the hypothesis stating the optimality of Gaussian QMLE is generally rejected. Finally, we consider non-parametric machine learning models forVaR estimation. These methods are designed to eliminate model risk because they are not based on a specific form of volatility. We use the support vector machine model for regression (SVR) based on the least square loss function (LS). In order to improve the solution of LS-SVR model, we used the weighted LS-SVR and the fixed size LS-SVR models. Numerical illustrations highlight the contribution of the proposed models for VaR estimation taking into account the risk of specification and estimation.

Keywords. Conditional VaR, Efficient estimator, Estimation risk, GARCH models, generalized Quasi maximum likelihood, Machine learning models, Misspecifi-

cation risk, Model risk.

Table of Contents

List of Tables

List of Figures

Chapitre 1

Introduction

Nous nous intéressons dans le cadre de cette thèse à la modélisation de la valeur à risque conditionnelle (notée VaR). Nous considérons le risque de modèle, qui englobe les risques de spécification et d'estimation. Le comité de BALE II a suggéré aux régulateurs des marchés financiers de prendre en compte le risque du modèle. Un modèle erroné pour l'estimation du risque peut en effet entraîner une sous-estimation ou une sur-estimation de la perte maximale pour un horizon et un niveau de confiance donnés. Par conséquent, il est crucial de tenir compte du problème de mauvaise spécification éventuelle du modèle, ainsi que du risque d'estimation. Souvent, des modèles paramétriques de variance-covariance sont utilisés dans ce cadre d'analyse de risque. Pour étudier le problème du risque d'estimation dans un cadre paramétrique, nous considérons dans le premier chapitre les modèles GARCH où nous développons une approche en deux étapes pour estimer la VaR conditionnelle. La première étape consiste à utiliser l'estimateur quasi-maximum de vraisemblance généralisé (gQMLE) pour estimer le paramètre de volatilité. La deuxième étape consiste à estimer le quantile théorique des innovations en utilisant le quantile empirique des résidus de la première étape. Nous considérons une densité instrumentale h pour l'estimateur gQML. Lorsque cette densité n'est pas la densité gaussienne, ou bien la vraie distribution des innovations, on obtient à la fois une estimation asymptotiquement biaisée de la volatilité et du quantile. Les deux erreurs se contrebalancent et nous obtenons finalement un estimateur convergent de la VaR conditionnelle. Nous avons obtenu la loi asymptotique du paramètre de VaR conditionnelle en utilisant le gQMLE pour une large classe de modèles GARCH. Dans ce chapitre, on a aussi proposé une méthode adaptative qui permet de déterminer la densité instrumentale h optimale. Nous avons obtenu des gains d'efficacité par rapport au QMLE fondé sur la densité gaussienne, particulièrement lorsque les innovations sont à queue lourde. De plus, nous avons étendu notre approche à l'estimation de mesures de distorsion de risque (DRM). Le deuxième chapitre propose un test permettant de déterminer si le QMLE le plus efficace est obtenu avec une densité différente de la gaussienne ou non. Plus précisément nous définissons h comme une distribution double Gamma généralisée (dGg), qui contient comme cas particulier la densité gaussienne. Cette loi contient trois paramètres, $h = dGg(b,p,d)$, mais nous trouvons que la variance asymptotique de l'estimateur de la VaR ne dépend que de d. Pour $d = 2$ on tombe sur le cas gaussien et un estimateur classique de la VaR conditionnelle, sinon pour des valeurs optimales de $d \neq 2$ notre estimateur fondé sur un gQMLE non gaussien est asymptotiquement plus efficace. Pour effectuer le test que la valeur optimale est $d = 2$ on estime dans une première étape le paramètre de VaR et on estime par la suite le paramètre qui minimise la variance asymptotique de notre estimateur. Dans une deuxième étape, nous définissons une statistique du test qui suit asymptotiquement un $\chi^2(1)$ lorsque la densité instrumentale optimale est gaussienne. Dans ces deux chapitres on a pu montrer que, dans le cadre de modèles GARCH, en utilisant un estimateur de la VaR fondé sur un gQMLE, nous pouvons réduire le risque d'estimation.

Dans le troisième chapitre nous utilisons des modèles d'apprentissage automatique pour la modélisation de la volatilité ainsi que de la VaR conditionnelle. L'avantage des méthodes utilisées, est qu'elles ne se basent pas sur des hypothèses tel que la normalité, et qu'elles ne nécessitent pas une forme explicite de la volatilité en fonction des rendements passés. Afin d'éviter le problème de spécification, nous proposons d'utiliser les modèles machine à vecteur de support ou les séparateurs à vaste marge (notés SVM). Généralement, dans un espace de dimension fini, la méthode SVM est considérée comme une méthode paramétrique de classification basée sur la fonction de perte Hinge. Cependant, au niveau d'un espace de dimension infinie, la méthode SVM sera considérée comme une méthode non-paramétrique de classification (SVC) ou de régression (SVR) basée sur différentes fonctions de perte. Dans notre cas, nous nous intéressons aux méthodes de régression. En se basant sur la fonction de perte moindres carrés (en anglais LS), la différence entre le SVR empirique et le SVR théorique vérifie la normalité asymptotique à vitesse $\sqrt{n}$. De ce fait, nous proposons d'utiliser la méthode LS-SVR et ses variantes pour estimer la VaR conditionnelle. En effet, pour améliorer la robustesse de la méthode LS-SVR, on propose d'utiliser la méthode de pondération notée (WLS-SVR). Puisque la solution de la méthode LS-SVR perd la propriété de parcimonie (sparsness), dans ce cas nous avons proposé d'utiliser la méthode de taille fixe notée (FS-LS-SVR). Notre étude a montré que la méthode LS-SVR est meilleure que la méthode générale SVR basée sur la fonction de perte "$\epsilon-$insensitive loss function" (ILF). La méthode SVR basée sur la fonction de perte ILF ne vérifie pas la propriété de la normalité asymptotique. D'autre part, les variantes de la méthode LS-SVR font mieux pour l'estimation de la VaR conditionnelle. Donc nous tirons la conclusion que les méthodes non-paramétriques peuvent être considérées comme des alternatives intéressantes pour évaluer la VaR conditionnelle. Nous terminons par une conclusion générale et des propositions de futurs axes de recherche.

1.1 Valeur à risque et méthodes de calcul

Afin de bien comprendre l'évolution des risques financiers et de sensibiliser les acteurs des marchés financiers aux différents risques, des mesures de risque ont été développées. L'accord de Bâle de 1988 impose une réglementation du risque visant à assurer la stabilité du système bancaire international en fixant une limite minimale à la quantité de fonds propres des banques. Par la suite JP-Morgan (1994) a lancé la méthodologie Riskmetrics, qui a permis une diffusion de nombreuses méthodes d'estimation de la valeur à risque (VaR). Dans le domaine de la gestion des risques, la mesure de risque est un outil à plusieurs vocations. De ce fait il est important de considérer la mesure de risque d'une part comme un outil réglementaire et d'autre part comme un outil stratégique d'aide à la prise de décision. Il est reconnu que la VaR est sans aucun doute l'outil le plus utilisé pour mesurer et contrôler les risques financiers. La VaR a été développée vers la fin des années 1980 pour répondre au besoin des institutions financière de calculer les allocations de fonds nécessaires pour faire face à des risques de marché dûs aux variations des prix. Jorion (1997) et Dowd (1998) ont été les premiers à mettre en évidence l'utilité de la VaR pour la quantification du risque. Par définition, la VaR est conçue comme une mesure synthétique de risque qui permet aux régulateurs des marchés financiers de déterminer, à un horizon temporel τ donné (annuel, mensuel ou journalier) et à un niveau de confiance α, la perte potentielle maximale qu'un portefeuille ou un actif peut subir sous des conditions normales de marché. Un des avantages de la VaR est qu'elle est facile à interpréter, et relativement facile à estimer. Ces avantages intéressent souvent les preneurs de risque afin d'améliorer leurs stratégies de prise de décision. Pour mieux comprendre le concept de la VaR on donne cet exemple illustratif.

Exemple 1.1 *Si la VaR d'un portefeuille à un niveau de confiance de 95% sur 10 jours ouvrables est de l'ordre de 1000 dollars, alors la perte maximale à un risque de 5 % ne dépasse pas les 1000 dollars sur l'horizon défini.*

Par définition, au niveau de risque $\alpha \in [0,1]$, la VaR (conditionnelle) de la série de rendements (ϵ_t) est l'opposée du quantile d'ordre α de la distribution conditionnelle :

$$\text{VaR}_t(\alpha) = -\inf\left\{x : P(\epsilon_{t+1} \leq x \mid \epsilon_u, u \leq t) \geq \alpha\right\}. \tag{1.1}$$

La figure 1.1 illustre un exemple de présentation de la VaR pour deux différentes fonctions de densité, une distribution normale et une distribution de Student avec 10 degrés de liberté (df=10). Nous constatons que la VaR de la distribution leptokurtique est plus large que celle de la distribution normale pour le même niveau de risque.

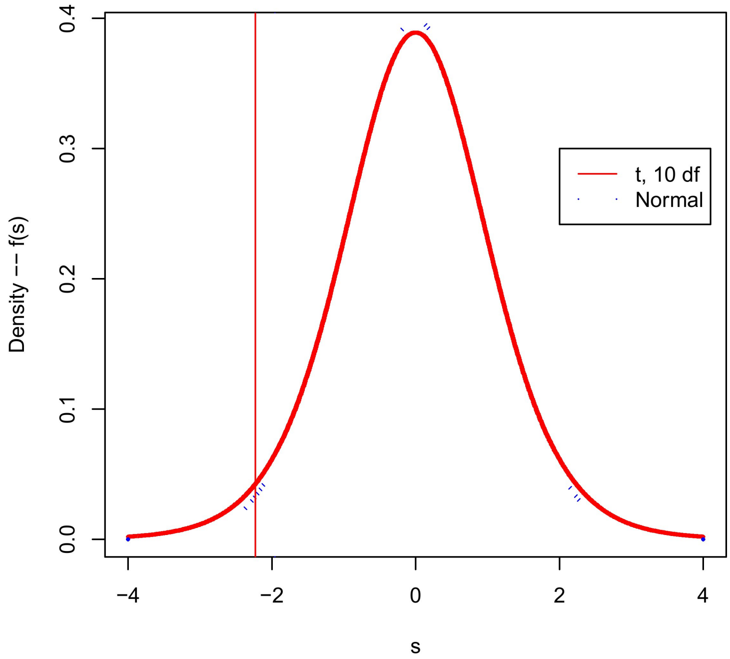


FIG. 1.1 – *Représentation graphique de la VaR pour la distribution de Student (df=10) et la distribution normale à un niveau de risque 5%.*

Il n'existe pas une méthode unique pour l'évaluation de la VaR. On distingue souvent trois classes: les approches non-paramétriques, semi-paramétriques et paramétriques. Les institutions financières ont intérêt à appliquer plusieurs méthodes d'évaluation de la VaR afin de minimiser le risque de se tromper lors de la détermination des fourchettes des pertes éventuelles.

i) Le principe des approches non-paramétrique est qu'elles ne supposent aucune distribution paramétrique de perte et profit. Parmi ces méthodes, on peut citer la méthode de simulation historique (HS), la méthode de simulation historique pondérée (WHS) et la méthode de simulation monte carlo (MCS). La méthode HS revient à calculer un fractile de la distribution empirique des rendements historiques, sans tenir compte de la dynamique des rendements. La méthode WHS attribue des pondérations aux rendements en fonction de leur ancienneté (aged WHS), en fonction de la volatilité des prix du marché (volatilité WHS) ou en fonction des corrélations passées et futures (corrélation WHS). Le principe de la méthode MCS est de faire un grand nombre de simulations de facteurs de risque, sous des hypothèses émises par l'investisseur, dans le cas où la densité de probabilité est inconnue ou difficile à déterminer. Enfin la VaR est estimée, comme pour la méthode HS, sur la base de la distribution empirique des pertes générées.

ii) La théorie des valeurs extrêmes (TVE) est souvent utilisée comme méthode semi-paramétrique de calcul de VaR. La TVE est différente des autres méthodes dans la mesure où elle s'intéresse uniquement aux queues de la distribution, au lieu de la modélisation totale de cette distribution. Il existe deux principaux types de modèles pour les valeurs extrêmes: l'approche des maxima par blocs, où l'échantillon peut être séparé en k blocs disjoints de même longueur et on se concentre sur la série des maxima dans ces k blocs, et l'approche d'excès au delà du seuil (POT) qui s'intéresse aux dépassements d'un seuil élevé et les modélisent séparément du reste des observations. Par conséquent, sur la base du théorème de Fisher et Tippet (1928), deux distributions majeures sont développées pour l'estimation du quantile. La distribution asymptotique d'une série de maximum est modélisée sous certaines conditions par la loi des valeurs extrêmes généralisées (GEV), tandis que la distribution pareto généralisée (GPD) est utilisée pour approximer la distribution des excès au dessus d'un certain seuil.

iii) Les méthodes paramétriques de la VaR reposent sur des hypothèses apparemment plus restrictives. Parmi les hypothèses utilisées la normalité des prix de marchés et la linéarité des espérances conditionnelles des profits. Parmi les méthodes paramétriques, on cite le modèle Riskmetrics développé par la banque JP-Morgan au début des années 90. Cette méthode se base sur l'estimation de la volatilité en accordant plus de poids aux rendements r_t les plus récents. Le modèle Riskmetrics se base sur l'hypothèse de normalité. Cette approche utilise la méthode de moyenne mobile pondérée (EWMA) afin de représenter la mémoire du marché. La volatilité dans ce cas est exprimée comme suit

$$\sigma^2_{t+1|t} = \frac{\sum_{\tau=0}^{\infty} \lambda^\tau r^2_{t-\tau}}{\sum_{\tau=0}^{\infty} \lambda^\tau}, \tag{1.2}$$

avec $0 < \lambda < 1$. Par la suite la VaR peut être obtenue en multipliant $\sigma_{t+1|t}$ par le quantile $(1-\alpha)$ de la distribution normale. Dans les modèles de type GARCH la volatilité est, comme dans (1.2), une moyenne pondérées des carrés des rendements passés, mais les poids sont choisis de manière plus souple. Pour l'estimation de la VaR, on remplace généralement le quantile de la loi normale par le quantile empirique des résidus du modèle GARCH, ce qui permet également de considérer cette méthode comme étant semi-paramétrique.

Example 1.1 *Étant donné une série de* 500 *observations, on estime la VaR à un seuil de confiance de 99% en utilisant les méthodes suivantes HS, WHS, Riskmetrics, GARCH et GARCH avec un bruit qui suit la distribution Student. Pour* *la Figure 1.2, on estime la VaR sur un échantillon de validation de taille* 250, *on remarque que les modèles HS et WHS donnent des estimations de la VaR approximativement semblables. De même pour les modèles paramétriques, on obtient des résultats différents de ceux obtenus par les méthodes non-paramétriques, mais proches entre eux.*

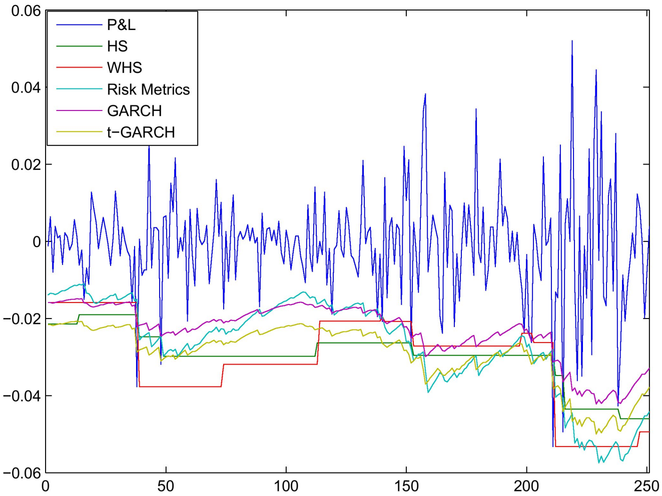


FIG. 1.2 – *Estimation de la VaR au niveau de risque 1% en utilisant les méthodes paramétriques et non-paramétriques.*

Suite à la différence entre les modèles utilisés pour estimer la VaR d'une même série de données, les institutions financières doivent déterminer les meilleurs modèles de prévision en se basant sur des backtests. La procédure de backtesting consiste à comparer les pertes prévues avec celles réelles. Depuis 1990 plusieurs méthodes de backtesting ont été proposées pour juger qu'un modèle de VaR est valide ou non. Les méthodes de backtesting peuvent être inconditionnelles ou conditionnelles. Les méthodes de couverture inconditionnelles proposées par Kupiec (1995) visent essentiellement à déterminer si la VaR est violée plus ou moins $\alpha.100\%$ fois. Si les exceptions sont dans des limites statistiques, le modèle est accepté, sinon il est rejeté. En revanche les méthodes de couverture conditionnelle ont pour objectif de tester si les exceptions sont indépendantes les une des autres. Christoffersen et Pelletier (2004) ont proposé cette méthode afin de régler les problèmes de couverture inconditionnelle non seulement en examinant la fréquence des violations de la VaR mais aussi le moment où elles se produisent. Engle et Manganelli (2004) ont proposé le test de quantile dynamique (DQ) qui consiste à tester l'hypothèse d'efficience conditionnelle. Ces derniers considèrent un modèle linéaire liant les violations courante à l'ensemble des violations passée et testent si les violations sont auto-corrélées. Cependant, la variable dépendante est une variable binaire. Étant donnée l'aspect dichotomique de la série des violations, Dumitrescu et al. (2012) ont suggéré d'étendre le test DQ à un test basé sur des modèles non-linéaires appelé le test dynamique binaire (DB). En effet le test DB consiste à utiliser le modèle dichotomique dynamique probit ou logit permettant de modéliser la dépendance non-linéaire entre les violations courantes et passées. Candelon et al. (2011) ont proposé le test "Duration-based GMM" qui se base sur l'approche GMM et les durées entre deux violations de la VaR. Le principe de ce test revient à vérifier si les violations de prévisions de la VaR se produisent de façon aléatoire.

1.2 Les risques liés à la modélisation de la VaR

Comme on l'a détaillé dans la section précédente, il y a plusieurs modèles d'estimation de la VaR. Face à ce problème le comité de Bâle II a recommandé aux institutions financières de prendre en considération le risque de modèle. En effet, il est envisageable que différents modèles produisent des résultats très différents sur une même base de données. De ce fait, la prise de décision sera erronée suite au mauvais choix de modèle.

Le risque de modèle englobe généralement deux grands types de risques: risque d'estimation et risque de spécification. En effet, les institutions financières ne connaissent généralement pas le processus générateur de données (DGP), ou même en admettant qu'il se trouve dans une certaine classe de modèles, il reste toujours une incertitude liée à l'estimation des paramètres, qui est fondée sur une information limitée et des hypothèses précises. Cependant, le risque de spécification est essentiellement dû au mauvais choix du modèle économétrique. Ceci soit au niveau du choix de l'ordre du modèle GARCH ou bien au niveau du choix du modèle GARCH symétrique ou asymétrique tel que les modèles GJR, EGARCH, TGARCH proposés par Glosten et al. (1993), Nelson (1991) et Zakoïan (1994). On détaillera plus ce concept au niveau du Chapitre 4.

Pour tenir compte des deux types de risque dans l'évaluation de la VaR, il est donc raisonnable de comparer plusieurs modèles paramétriques (par exemples quelques-uns des nombreux modèles de type GARCH), et de comparer plusieurs méthodes d'estimation de ces modèles (par exemple plusieurs gQMLE), ou bien d'utiliser des méthodes non-paramétriques d'apprentissage automatique qui reposent sur des hypothèses moins restrictives.

1.2.1 Modèles paramétriques

Les modèles d'estimation de la volatilité souvent utilisés sont les modèles ARCH et leur généralisation GARCH proposés par Engle (1982) et Bollerslev (1986) respectivement. En effet, pour une séquence (ϵ_t) des rendements, un modèle hétéroscedastique est généralement de la forme suivante

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t = \sigma_t(\theta_0) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots ; \theta_0) \end{cases} \tag{1.3}$$

où (η_t) est une suite indépendamment et identiquement distribuée (iid) centrée, $\theta_0 \in \mathbb{R}^m$ est un paramètre de l'espace compact Θ.

Prenons le cas où (ϵ_t) pour $t \in \mathbb{Z}$ est un processus GARCH(p,q) vérifiant la forme suivante

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \end{cases} \tag{1.4}$$

où les α_i,β_j et ω sont des constantes positives. L'estimation du modèle (1.4) revient à estimer le vecteur de paramètre $\theta = (\omega, \alpha_1, \ldots, \beta_1, \ldots)' \in]0,\infty[* [0,\infty[^{(p+q)}$ en utilisant la méthode du maximum de vraisemblance (MV) ou quasi-maximum de vraisemblance (QMV, en anglais QML). Il est à noter que la vraie valeur de θ est inconnue.

On note aussi que (η_t) est une suite iid centrée de variance 1 et de densité f inconnue.

L'estimateur le plus utilisé des modèles de type ARCH est sans doute le QMLE gaussien. La consistance et la normalité asymptotique (CAN) de cet estimateur ne nécessitent que quelques hypothèses de régularité et la condition standard d'identifiabilité $E\eta_t^2 = 1$ (voir Berkes, Horváth et Kokoszak (2003) et Francq et Zakoïan (2004) pour le cas des modèles GARCH standard, Mikosch et Straumann (2006), Bardet et Wintenberger (2009) pour des modèles généraux). Dans le cadre du modèle GARCH standard, Berkes et Horváth (2004) ont introduit le QMLE non Gaussien et ont établi leur CAN sous des conditions d'identifiabilité alternatives. Pour le modèle général (1.3), Francq et Zakoïan (2004) ont montré que des QMLE généralisés particuliers conduisent à des prédictions en une seule étape des puissances $|\epsilon_t|^r$ pour $r \in \mathbb{R}$. Dans le cas de modèle GARCH(p,q) standard Francq et al. (2011) ont montré que, sous l'hypothèse d'identifiabilité classique $E\eta_t^2 = 1$ pour le modèle d'origine et $E|\eta_t^{(r)}|^r = 1$ pour le modèle reparamétré, le QMLE généralisé est plus efficace que le QMLE gaussien. Francq et Zakoïan (2013) ont démontré que le QMLE peut être consistant sous cette condition d'identifiabilité pour une classe de densité instrumentale particulière tel que la densité gaussienne généralisée.

1.2.2 Modèles non-paramétriques d'apprentissage automatique

La régression non paramétrique est une forme d'analyse dans laquelle le prédicteur ne prend pas une forme prédéterminée, mais il est construit selon les informations issues des données. L'objectif de la régression non paramétrique est d'estimer la fonction de régression directement, plutôt que d'estimer les paramètres. Ces modèles servent à résoudre le problème de régression paramétrique qui conduit souvent à des résultats erronés dûs à la non-concordance entre la structure de modèle et le processus générateur des données réelles.

La régression non-paramétrique consiste à estimer une relation fonctionnelle entre la variable indépendante $X \in \mathbb{R}^d$ et la variable dépendante $Y \in \mathbb{R}$ sans hypothèse précise sur la distribution conjointe de (X,Y). Le but est de construire un "bon prédicteur" $f : \mathbb{R}^d \to \mathbb{R}$. Nous entendons par bon prédicteur que f minimise la perte attendue, *i.e.* le risque

$$R_{L,P}(f) = E_P \left[L(X,Y,f(X)) \right],$$

où P représente la distribution jointe inconnue de (X,Y) et $L : X * Y * \mathbb{R} \to [0,\infty)$ est la fonction de perte. Nous prenons dans ce qui suit le cas de la fonction de perte moindre carrée (LS en anglais) définie par

$$L_{LS}(y) = (f(x) - y)^2. \tag{1.5}$$

Selon Steinwart et Christmann (2008) et Gyorfi et al. (2002), la fonction de perte LS est généralement utilisée puisqu'elle conduit naturellement à des estimations qui peuvent être calculées rapidement.

En outre, parmi les modèles de régression non-paramétrique, nous utilisons dans cette thèse le modèle SVR développé par Vapnik (1998). Ce modèle non-paramétrique ne suppose aucune forme de la fonction de régression ni de la distribution des résidus.

La méthode SVR repose essentiellement sur une fonction de perte qui peut prendre plusieurs formes telles que la fonction $\epsilon -$ insensitive. Selon Hable (2012), cette fonction de perte n'est pas deux fois différentiable. En effet pour éviter ce problème, Hable (2012) a mentionné qu'en utilisant la fonction de perte donnée par l'équation (1.5), la différence entre le SVR empirique et l'équivalent théorique est asymptotiquement normale. En raison de cette propriété, on a choisi d'utiliser comme méthode de régression non-paramétrique la méthode SVR basée sur la fonction de perte LS et ses variants pour améliorer la solution de cette dernière. On donnera plus de détails concernant les méthodes non-paramétriques au chapitre 4.

1.3 Résultats du chapitre 2

Pour estimer le paramètre de la VaR conditionnelle (comme défini dans Francq et Zakoïan, 2012), une méthode naturelle consiste à estimer le paramètre de volatilité par QMLE gaussien, puis à estimer le quantile des innovations par le quantile des résidus du QMLE. Dans ce chapitre on vise à étendre cette approche en utilisant un gQMLE fondé sur une densité instrumentale h. La VaR conditionnelle de la série de rendement (ϵ_t) peut s'écrire de la manière suivante

$$\mathrm{VaR}_t(\alpha) = -\sigma_t(\theta_0)\xi_\alpha, \tag{1.6}$$

où ξ_α est le quantile théorique d'ordre α de la distribution des innovations. Notre approche consiste en premier lieu à estimer le paramètre du modèle (1.3) par gQML. En deuxième lieu on estime le quantile des innovations par le quantile empirique des résidus. Nous dérivons tout d'abord la consistance et la normalité asymptotique (CAN) du gQMLE pour le modèle (1.3) sous des hypothèses un peu similaires à celles proposées par Berkes et Horváth (2004) et Francq et Zakoïan (2012). Il est à noter que pour identifier le modèle (1.3) la contrainte standard est $E\eta_t^2 = 1$, mais dans ce qui suit nous n'aurons pas recours à cette hypothèse. Le gQMLE converge vers un paramètre $\theta_0^* \neq \theta_0$ et le quantile empirique des résidus converge vers $\xi_\alpha^* \neq \xi$. Néanmoins nous aurons généralement la consistance de l'estimateur de la VaR. On donnera plus de détails au niveau du Lemme 1.1. Nous aurons besoin de la condition suivante, qui est satisfaite pour tous les modèles de volatilité standard.

A1: Il existe une fonction continue H tel que pour tout $\theta \in \Theta$, pour tout $K > 0$, et pour toute suite $(x_i)_i$

$$K\sigma(x_1,x_2,\ldots;\theta) = \sigma(x_1,x_2,\ldots;H(\theta,K)).$$

Cette condition signifie que la forme paramétrique de la volatilité est stable par changement d'échelle. On peut alors définir le paramètre de VaR comme étant $\theta_{0,\alpha} = H(\theta_0, -\xi_\alpha) = H(\theta_0^*, -\xi_\alpha^*)$. Nous allons obtenir, dans le théorème 1.1 qui suit, la loi asymptotique de $\hat{\theta}_{n,\alpha} - \theta_{0,\alpha}^*$. Ensuite, on prendra le cas du modèle standard (1.4) où on montrera que le choix de la densité instrumentale optimale h ne dépend que des propriétés de la distribution des η_t. On a développé une approche adaptative pour déterminer h optimale. En plus, on a étendu notre méthode en deux étapes à l'estimation du paramètre de risque à des mesures du risque de distorsion (DRM) satisfaisant la propriété d'additivité. Enfin, nous fournissons des illustrations numériques qui mettent en valeur l'optimalité du gQMLE par rapport au QMLE gaussien. Nous commençons tout d'abord par présenter le gQMLE.

1.3.1 Estimation de la VaR conditionnelle par gQMLE

Dans cette section on montre que le gQMLE de la VaR donne un estimateur consistant de

$$\mathrm{VaR}_t(\alpha) = -\sigma_{t+1}(\theta_0^*)\xi_\alpha^*, \tag{1.7}$$

où ξ_α^* désigne le quantile d'ordre α de $\eta_t^* := \eta_t/\sigma_*$, avec $\sigma^* > 0$ et

$$\theta_0^* = H(\theta_0,\sigma_*). \tag{1.8}$$

Par conséquent, le gQMLE de la volatilité converge vers $\sigma_t(\theta_0^*) = \sigma^*\sigma_t(\theta_0)$ et le quantile empirique des résidus de gQMLE converge vers $\xi_\alpha^* = \xi_\alpha/\sigma^*$.

Estimation du paramètre de volatilité

Étant donné $\epsilon_1,\dots,\epsilon_n$ une réalisation de longueur (n) du processus (ϵ_t), et des valeurs initiales arbitraires $\widetilde{\epsilon}_i$ pour $i \leq 0$, soit

$$\widetilde{\sigma}_t(\theta) = \sigma(\epsilon_{t-1},\epsilon_{t-2},\dots,\epsilon_1,\widetilde{\epsilon}_0,\widetilde{\epsilon}_{-1},\dots;\theta),$$

un proxy de

$$\sigma_t(\theta) = \sigma(\epsilon_{t-1},\epsilon_{t-2},\dots,\epsilon_1,\epsilon_0,\epsilon_{-1},\dots;\theta).$$

Pour $h > 0$, nous considérons le critère du QML,

$$\widetilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(\epsilon_t,\widetilde{\sigma}_t(\theta)), \qquad g(x,\sigma) = \log\frac{1}{\sigma}h\left(\frac{x}{\sigma}\right), \tag{1.9}$$

et le gQMLE

$$\hat{\theta}_n^* = \arg\max_{\theta\in\Theta}\widetilde{Q}_n(\theta).$$

On note que pour $h = \phi$ on obtient le QMLE Gaussien. On établit la CAN de $\hat{\theta}_n^*$ sous les hypothèses **A2-A12** données dans le chapitre 2 .

Notre hypothèse **A4** est moins restrictive que celle proposée par Francq et Zakoïan (2012) (cette référence sera notée FZ par la suite). En effet on n'a pas besoin de supposer qu'on obtient le maximum pour $\sigma_* = 1$. En plus on n'a pas de condition d'identifiabilité pour η_t (telle que $E\eta_t^2 = 1$). Nous avons seulement besoin de supposer l'existence de σ_* tel que

$$E\left\{\frac{\eta_0}{\sigma_*}\frac{h'}{h}\left(\frac{\eta_0}{\sigma_*}\right)\right\} = -1. \tag{1.10}$$

Le Lemme suivant étend les résultats obtenus par Berkes et Horváth (2004) pour le cas du GARCH standard, et par FZ sous d'autres hypothèses.

Lemme 1.1 (Comportement asymptotique du QMLE généralisé) *Si* **A1-A7** *sont satisfaites, alors*

$$\hat{\theta}_n^* \to \theta_0^*, \quad p.s.$$

où θ_0^* *est définit par (1.8). Si, en plus,* **A8-A12** *sont satisfaites et* $Eg_2(\eta_0,1) \neq 0$ *alors*

$$\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\tau_h J_*^{-1})$$

où

$$J_* = 4ED_t(\theta_0^*)D_t'(\theta_0^*) \quad et \quad \tau_h = \frac{4Eg_1^2(\sigma_*^{-1}\eta_0,1)}{\left\{Eg_2(\sigma_*^{-1}\eta_0,1)\right\}^2}, \tag{1.11}$$

avec

$$D_t(\theta) = \frac{1}{\sigma_t(\theta)}\frac{\partial \sigma_t(\theta)}{\partial \theta}, \quad g_1(x,\sigma) = \frac{\partial g(x,\sigma)}{\partial \sigma} \qquad et \qquad g_2(x,\sigma) = \frac{\partial g_1(x,\sigma)}{\partial \sigma}.$$

Estimation du paramètre de VaR

Pour le modèle de volatilité général (1.4), on a la forme de la VaR donnée par (1.7). Notons que si $\xi_\alpha^* < 0$, on a alors sous **A1**

$$\mathrm{VaR}_t(\alpha) = \sigma_{t+1}\left(\theta_{0,\alpha}\right) \quad \text{où} \quad \theta_{0,\alpha} = H\left(\theta_0^*, -\xi_\alpha^*\right).$$

Nous désignons par $\theta_{0,\alpha}$ le paramètre de VaR proposé par Francq et Zakoïan (2012). Notons que $\xi_\alpha := \sigma_* \xi_\alpha^*$ est le quantile d'ordre α de η_t. On a alors

$$\theta_{0,\alpha} = H\left(\theta_0^*, -\xi_\alpha^*\right) = H\left(\theta_0, -\xi_\alpha\right).$$

Soit $\hat{\xi}_{\alpha,n}^*$ le quantile empirique des résidus $\hat{\eta}_t^* := \epsilon_t/\widetilde{\sigma}_t(\hat{\theta}_n^*)$ pour $t = 1,\ldots,n$. Pour obtenir la distribution asymptotique de l'estimateur en deux étapes du paramètre VaR, on établit les résultats intermédiaires suivants.

Théorème 1.1 *On suppose que* η_1 *a une densité* f*, continue en* ξ_α*, telle que* $f(\xi_\alpha) > 0$*. Sous les hypothèses du Lemme 1.1, et en supposant que* **A5** *et* **A12** *sont vérifiées pour un* $\delta > 1$*, on aura*

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_n^* - \theta_0^* \\ \hat{\xi}_{\alpha,n}^* - \xi_\alpha^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0,\Sigma^* := \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\prime} & \Sigma_{22}^* \end{pmatrix}\right\},$$

où

$$\begin{aligned}
\Sigma_{11}^* &= \tau_h J_*^{-1}, \\
\Sigma_{12}^* &= -\left\{\xi_\alpha^* \tau_h - \frac{4c_\alpha}{\sigma_* f(\xi_\alpha) E g_2(\eta_0^*,1)}\right\} J_*^{-1}\Omega_*, \\
\Sigma_{22}^* &= \frac{\tau_h (\xi_\alpha^*)^2}{4} - \frac{2c_\alpha \xi_\alpha^*}{\sigma_* f(\xi_\alpha) E g_2(\eta_0^*,1)} + \frac{\alpha(1-\alpha)}{\sigma_*^2 f^2(\xi_\alpha)},
\end{aligned}$$

avec $\Omega_* = ED_t(\theta_0^*)$*,* $c_\alpha = Cov(\mathbf{1}_{\{\eta_t^* < \xi_\alpha^*\}}, g_1(\eta_t^*,1))$*.*

En appliquant la méthode delta on obtient les résultats suivants

Corollaire 1.1 *Sous les hypothèses du Théorème 1.1 et si h est différentiable aux points $(\theta_0^*, -\xi_\alpha^*)$, avec $\xi_\alpha^* < 0$, on obtient*

$$\sqrt{n}\left(\hat{\theta}_{n,\alpha}^* - \theta_{0,\alpha}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, G_* \Sigma^* G_*'\right),$$

où

$$G_* = \left[\frac{\partial H(\theta,K)}{\partial(\theta',K)}\right]_{(\theta_0^*, -\xi_\alpha^*)}.$$

Conformément à Francq et Zakoïan (2012), en utilisant le Corollaire 1.1 et la méthode delta, on peut obtenir des intervalles de confiance pour $\mathrm{VaR}_t(\alpha) = \sigma_{t+1}(\theta_{0,\alpha})$ pour un niveau de risque donné.

En se basant sur le Corollaire 1.2 on montre que l'estimateur du paramètre de VaR est invariant par changement d'échelle de la densité instrumentale.

Corollaire 1.2 *Sous les hypothèses du Corollaire 1.1, et si* **A1** *est vraie lorsqu'on remplace σ_t par $\tilde{\sigma}_t$,* i.e. *si*

$$K\tilde{\sigma}_t(\theta) = \tilde{\sigma}_t(\theta)\left\{H(\theta,K)\right\}, \tag{1.12}$$

alors l'estimateur $\hat{\theta}_{n,\alpha}^*$ ne change pas si $h(x)$ est remplacé par $h_s(x) = s^{-1}h(s^{-1}x)$, pour tout $s > 0$.

Dans ce chapitre, nous vérifions la régularité des conditions du Lemme 1.1 pour le modèle GARCH(1,1) standard, puis pour les modèles APARCH.

Cas du modèle GARCH(1,1) standard

Dans le cas standard des modèles GARCH, plusieurs hypothèses peuvent être rendues plus explicites. Soit la vraie valeur $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \beta_{0p})'$ et l'élément générique de Θ de la forme $\theta = (\omega, \alpha_1, \ldots, \beta_p)'$.

Une condition nécessaire et suffisante pour qu'une solution strictement stationnaire de (1.4) existe est $\gamma < 0$, où γ est l'exposant de Lyapunov maximal du modèle (voir e.g. Francq et Zakoïan (2004)). On note que $\gamma = \gamma(\theta_0)$ dépend de θ_0 (et aussi de la loi de η_1). Soit $\mathcal{A}_\theta(z) = \sum_{i=1}^{q} \alpha_i z^i$ et $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$. Dans ce cadre GARCH, les hypothèses **A2**, **A3**, **A6**, **A9**, **A10** et **A12** sont équivalentes à:

C: $\gamma(\theta_0) < 0$; $\forall \theta \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$ et $\omega > \underline{\omega}$ pour certaines $\underline{\omega} > 0$; $|\eta_0|$ a une distribution non dégénérée; si $p > 0$, $\mathcal{A}_{\theta_0}(z)$ et $\mathcal{B}_{\theta_0}(z)$ n'ont pas de racine commune, $\mathcal{A}_{\theta_0}(1) \neq 0$, et $\alpha_{0q} + \beta_{0p} \neq 0$.

Dans le cas du modèle GARCH(1,1), sous l'hypothèse **C**, la matrice G_* du Corollaire 1.1 est donnée par

$$G_* = \begin{pmatrix} (\xi_\alpha^*)^2 & 0 & 0 & -2\xi_\alpha^*\omega_0^* \\ 0 & (\xi_\alpha^*)^2 & 0 & -2\xi_\alpha^*\alpha_0^* \\ 0 & 0 & 1 & 0 \end{pmatrix} := \begin{pmatrix} A_* & -2\xi_\alpha^* \begin{pmatrix} \omega_0^* \\ \alpha_0^* \\ 0 \end{pmatrix} \end{pmatrix}.$$

Notons aussi, pour tout $\theta_0^* = (\omega_0^*,\alpha_0^*,\beta_0^*) \in \Theta$, on a

$$\begin{aligned}(\omega_0^*,\alpha_0^*,0)\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} &= \omega_0^* + \alpha_0^*\epsilon_{t-1}^2 + \beta_0^*\left\{(\omega_0^*,\alpha_0^*,0)\frac{\partial\sigma_{t-1}^2(\theta_0^*)}{\partial\theta}\right\} \\ &= \sum_{i=0}^{\infty}\beta_0^{*i}\left\{\omega_0^* + \alpha_0^*\epsilon_{t-i}^2\right\} = \sigma_t^2(\theta_0^*).\end{aligned}$$

Il s'ensuit que

$$\frac{1}{\sigma_t(\theta_0^*)}\frac{\partial\sigma_t(\theta_0^*)}{\partial\theta'}\begin{pmatrix}\omega_0^*\\ \alpha_0^*\\ 0\end{pmatrix} = \frac{1}{2} \quad p.s.,$$

et

$$\Omega_*'\begin{pmatrix}\omega_0^*\\ \alpha_0^*\\ 0\end{pmatrix} = \frac{1}{2}, \quad J_*\begin{pmatrix}\omega_0^*\\ \alpha_0^*\\ 0\end{pmatrix} = 2\Omega_*, \quad J_*^{-1}\Omega_* = \frac{1}{2}\begin{pmatrix}\omega_0^*\\ \alpha_0^*\\ 0\end{pmatrix}, \quad \Omega_*'J_*^{-1}\Omega_* = \frac{1}{4}.$$

La deuxième égalité de la ligne précédente montre que

$$\mathrm{Var}\left(\frac{1}{\sigma_t^2(\theta_0^*)}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta}\right) = J_* - 4\Omega_*\Omega_*' = J_*(J_*^{-1} - \Psi_*)J_*,$$

où

$$\Psi_* = \begin{pmatrix}\omega_0^*\\ \alpha_0^*\\ 0\end{pmatrix}\begin{pmatrix}\omega_0^* & \alpha_0^* & 0\end{pmatrix} = \begin{pmatrix}\omega_0^{*2} & \omega_0^*\alpha_0^* & 0\\ \omega_0^*\alpha_0^* & \alpha_0^{*2} & 0\\ 0 & 0 & 0\end{pmatrix}.$$

Sous **A9**, la matrice $\mathrm{Var}\left(\frac{1}{\sigma_t^2(\theta_0^*)}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta}\right)$ est définie positive. Il s'ensuit que

$$J_*^{-1} - \Psi_* \quad \text{est définie positive.} \tag{1.13}$$

En outre, nous avons

$$\begin{aligned}G_*\Sigma^*G_*{}' &= \tau_h A_* J_*^{-1}A_* + \left(\frac{4(\xi_\alpha^*)^2\alpha(1-\alpha)}{\sigma_*^2 f^2(\xi_\alpha)} - \tau_h(\xi_\alpha^*)^4\right)\Psi_* \\ &= \tau_h A_*(J_*^{-1} - \Psi_*)A_* + \frac{4(\xi_\alpha^*)^2\alpha(1-\alpha)}{\sigma_*^2 f^2(\xi_\alpha)}\Psi_*.\end{aligned}$$

Pour la dernière égalité, nous avons utilisé $A_*\Psi_*A_* = (\xi_\alpha^*)^4\Psi_*$.

Maintenant, nous introduisons les analogues de certain symboles étoilés, qui sont indépendants de la densité instrumentale h, en utilisant la matrice de transformation

$$M_* = \begin{pmatrix}\frac{1}{\sigma_*^2}I_2 & 0_2\\ 0_2' & 1\end{pmatrix}.$$

Nous définissons ainsi $A = M_*^{-1} A_*$ et $\Psi = M_* \Psi_* M_* = \sigma_*{}^{-4} \Psi_*$. Notons aussi que

$$\theta_0 = M_* \theta_0^*, \quad D_t(\theta_0^*) = M_* D_t(\theta_0) \quad \text{et} \quad J_* = M_* J M_*.$$

Avec ces notations, nous avons

$$G_* \Sigma^* G_*{}' = \tau_h A (J^{-1} - \Psi) A + \frac{4\xi_\alpha^2 \alpha(1-\alpha)}{f^2(\xi_\alpha)} \Psi. \tag{1.14}$$

La densité instrumentale h_1 est dite plus efficace que h_2, ce que l'on note $h_1 \succ h_2$, si la différence entre les variances asymptotiques données par l'équation (1.14) est définie positive. Pour la variance asymptotique, uniquement τ_h dépend de h. Compte tenu de l'équation (1.13), ceci montre que $h_1 \succ h_2$ si et seulement si $\tau_{h_1} < \tau_{h_2}$.

Cas du modèle APARCH

Le modèle APARCH a été introduit par Ding, Granger et Engle (1993). Ce modèle inclut le cas du modèle GARCH standard, le modèle TARCH (Zakoïan 1994), le modèle GJR de Glosten, Jagannathan et Runkle (1993) parmi d'autres. Soient $x^+ = \max(x,0)$ et $x^- = \max(-x,0)$, le modèle APARCH se définit comme suit

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega_0 + \sum_{i=1}^q \alpha_{0i+} (\epsilon_{t-i}^+)^\delta + \alpha_{0i-} (\epsilon_{t-i}^-)^\delta + \sum_{j=1}^p \beta_{0j} \sigma_{t-j}^\delta \end{cases} \tag{1.15}$$

où les coefficients satisfont $\alpha_{0i+} \geq 0$, $\alpha_{0i-} \geq 0$, $\beta_{0j} \geq 0$, $\omega_0 > 0$ et $\delta > 0$. Le modèle GARCH standard est obtenu pour $\delta = 2$ et $\alpha_{0i-} = \alpha_{0i+}$. Selon Hamadeh et Zakoïan (2011) le paramètre de puissance δ n'est plus facile à estimer, de ce fait nous considérons δ fixe.

Comme dans l'hypothèse **C**, soit $\gamma(\theta_0)$ l'exposant de Lyapunov maximal associé au modèle (1.15). Hamadeh et Zakoïan (2011) ont montré la CAN du QMLE gaussien de $\theta_0 = (\omega_0, \alpha_{01+}, \ldots, \alpha_{0q-}, \beta_{01}, \ldots, \beta_{0p})'$ sous l'hypothèse:

D: $\gamma(\theta_0) < 0$; θ_0 appartient à l'intérieur de Θ; il existe $\underline{\omega} > 0$ tel que, $\forall \theta \in \Theta$, $\omega > \underline{\omega}$ et $\sum_{j=1}^p \beta_j < 1$; le support de la distribution de η_1 contient au moins trois points; $P[\eta_t > 0] \in (0,1)$; si $p > 0$, $\mathcal{B}_{\theta_0}(z)$ n'a pas de racine commune avec $\mathcal{A}_{\theta_0+}(z) = 1 - \sum_{i=1}^q \alpha_{0i+} z^i$ et $\mathcal{A}_{\theta_0-}(z) = 1 - \sum_{i=1}^q \alpha_{0i-} z^i$; $\mathcal{A}_{\theta_0+}(1) + \mathcal{A}_{\theta_0-}(1) \neq 0$ et $\alpha_{0q,+} + \alpha_{0q,-} + \beta_{0p} \neq 0$ (avec la notation $\alpha_{00,+} = \alpha_{00,-} = \beta_{00} = 1$)

et sous la condition d'identifiabilité $E\eta_1^2 = 1$.

Le théorème suivant étend les résultats obtenus dans la section précédente.

Théorème 1.2 *Considérons le modèle APARCH(p,q) (1.15) sous l'hypothèse* **D**. *Supposons que* η_1 *a une densité* f, *continue en* $\xi_\alpha < 0$, *tel que* $f(\xi_\alpha) > 0$. *Si la* *densité instrumentale h satisfait* **A4**, **A5**, **A7**, **A8** *et* **A11**, *alors l'estimateur en deux étapes du paramètre de VaR au niveau de confiance $\alpha \in (0,1)$ satisfait*

$$\sqrt{n}\left\{\hat{\theta}^*_{n,\alpha} - H\left(\theta^*_0, -\xi^*_\alpha\right)\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, G_*\Sigma^* G_*'\right),$$

où, pour $\xi > 0$,

$$H\left(\omega, \alpha_{1+}, \ldots, \alpha_{q-}, \beta_1, \ldots, \beta_p, \xi\right) = \left(\xi^\delta \omega, \xi^\delta \alpha_{1+}, \ldots, \xi^\delta \alpha_{q-}, \beta_1, \ldots, \beta_p\right)$$

et

$$G_*\Sigma^* G_*' = \tau_h A(J^{-1} - \Psi)A + \frac{4\xi^2_\alpha \alpha(1-\alpha)}{f^2(\xi_\alpha)}\Psi,$$

où $\overline{\theta}_0' = (\omega_0, \alpha_{01+}, \ldots, \alpha_{0q-}, 0, \ldots, 0)$,

$$A = \mathrm{diag}\left\{(-\xi_\alpha)^\delta I_{2q+1}, I_p\right\}, \quad \Psi = \overline{\theta}_0 \overline{\theta}_0', \quad J = 4ED_1(\theta_0)D_1'(\theta_0).$$

Pour les densités instrumentales h_1 et h_2 , nous avons $h_1 \succ h_2$ si et seulement si $\tau_{h_1} < \tau_{h_2}$.
Ce théorème prouve qu'une densité instrumentale h ayant le τ_h minimal est optimale. En particulier l'optimalité de h: 1) ne depend pas de θ_0^*, ni du modèle de volatilité; 2) ne dépend pas de α.

Choix optimal de la densité instrumentale

Compte tenu du Théorème 1.2, la densité instrumentale optimale h, appartenant à une classe de densités satisfaisant les hypothèses du théorème, est celle qui a le τ_h le plus petit. Dans notre étude nous avons considéré deux exemples de densité instrumentale h, la GED et la Student. Pour la première distribution, τ_h s'exprime en fonction de moments qui peuvent être estimés empiriquement. Pour le cas de la distribution Student, τ_h n'a pas de forme explicite, mais il peut aussi être facilement estimé.

Considérons le cas où h est une GED(κ). La valeur κ_0 de κ qui minimise

$$\tau_{GED} = \frac{4}{\kappa^2}\left(\frac{E\left|\eta_1\right|^{2\kappa}}{\left(E\left|\eta_1\right|^\kappa\right)^2} - 1\right), \tag{1.16}$$

est le paramètre optimal de la GED. Nous pouvons dire alors que la GED(κ_0) est une GED-optimale. L'estimateur empirique de κ_0 peut être obtenu comme suit. Soient $\hat{\eta}_t = \epsilon_t / \tilde{\sigma}_t(\hat{\theta}_n)$, $t = 1, \ldots, n$, les résidus obtenus à partir d'une procédure d'estimation de première étape, qui est consistante mais pas nécessairement optimale, par exemple celle du QMLE gaussien. Un estimateur de κ_0 est alors défini par

$$\hat{\kappa} = \arg\min_{\kappa \in \mathcal{K}} \frac{1}{\kappa^2}\left(\frac{\hat{\mu}_{2\kappa}}{\hat{\mu}^2_\kappa} - 1\right), \quad \hat{\mu}_r = \frac{1}{n}\sum_{t=1}^n |\hat{\eta}_t|^r,$$

où $\mathcal{K}$ est un intervalle borné contenant κ_0.

Considérons maintenant le cas de la densité instrumentale Student(ν), avec ν le degré de liberté. Les paramètres σ_* et τ_h peuvent être estimés comme suit. Soient $\hat{\eta}_1,\dots,\hat{\eta}_n$ les résidus d'une procédure d'estimation de première étape. Soient C et S des sous-ensembles compacts de $]0,\infty[$. Pour toute valeur de $\nu \in C$, σ_* peut être estimé par

$$\hat{\sigma}_* = \arg\max_{\sigma \in S} \sum_{t=1}^{n} g(\hat{\eta}_t,\sigma).$$

Nous obtenons alors un estimateur du paramètre de la densité instrumentale "Student-optimal" comme suit

$$\hat{\nu} = \arg\min_{\nu \in C} \frac{n^{-1}\sum_{t=1}^{n} g_1^2\left(\hat{\sigma}_*^{-1}\hat{\eta}_t,1\right)}{\left\{n^{-1}\sum_{t=1}^{n} g_2\left(\hat{\sigma}_*^{-1}\hat{\eta}_t,1\right)\right\}^2}. \tag{1.17}$$

Sous optimalité de l'approche adaptative

Supposons une forme paramétrique de $h_\kappa(x)$, $\kappa \in \mathcal{K}$ pour la densité instrumentale. Nous savons que la densité instrumentale optimale est la distribution (inconnue) f de η_1, ou de façon équivalente une version mise à l'échelle $\sigma^{-1}f(x/\sigma)$, $\sigma > 0$, de cette densité (voir Corollaire 1.2). Si une versions mise à l'échelle de f appartient à la classe paramétrique des densités instrumentales, *i.e.* si $f(x) = \sigma_0^{-1}h_{\kappa_0}(x/\sigma_0)$ pour un $\kappa_0 \in \mathcal{K}$ et un $\sigma_0 > 0$, alors la densité instrumentale optimale peut être trouvée par la procédure du (quasi-) maximum de vraisemblance

$$(\hat{\kappa},\hat{\sigma}) = \arg\max_{(\kappa,\sigma)\in\mathcal{K}\times(0,\infty)} \sum_{t=1}^{n} \log \sigma^{-1}h_\kappa(\hat{\eta}_t/\sigma),$$

où $\hat{\eta}_t = \epsilon_t/\tilde{\sigma}_t(\hat{\theta}_n)$, $t = 1,\dots,n$, sont les résidus obtenus à partir du QMLE gaussien ou de toute autre procédure d'estimation consistante de première étape.

Même si f n'appartient pas à la classe des densités, la procédure converge dans des conditions de régularité générale (voir White 1982) vers un minimiseur d'une divergence de Kullback-Leibler tel que

$$(\kappa^*,\sigma^*) = \arg\max_{(\kappa,\sigma)\in\mathcal{K}\times(0,\infty)} E\log\sigma^{-1}h_\kappa(\eta_1/\sigma).$$

Considérons le cas où h appartient à la classe de distributions GED(κ). Nous avons alors

$$\sigma^* = \left(\frac{\kappa^* E|\eta_1|^{\kappa^*}}{2}\right)^{1/\kappa^*},$$

où

$$\kappa^* = \arg\max_{\kappa\in\mathcal{K}} \log\left(\frac{\kappa}{\Gamma(1/\kappa)2^{1+1/\kappa}}\right) - \frac{1}{\kappa}\left\{\log\left(\frac{\kappa E|\eta_1|^\kappa}{2}\right)+1\right\}.$$

Soit GED(κ_0) la densité instrumentale GED-optimale, et soit τ_0, la valeur de τ_h correspondant à cette densité. Étant donné (1.16), nous avons

$$\tau_0 = \frac{4}{\kappa_0^2}\left(\frac{E\,|\eta_1|^{2\kappa_0}}{(E\,|\eta_1|^{\kappa_0})^2} - 1\right), \quad \kappa_0 = \arg\min_{\kappa} \frac{4}{\kappa^2}\left(\frac{E\,|\eta_1|^{2\kappa}}{(E\,|\eta_1|^{\kappa})^2} - 1\right).$$

Soit τ^*, la valeur de τ_h lorsque h est la GED(κ^*). Ce τ^* est optimal lorsque la densité f de η_1 est une GED mise à l'échelle, et dans ce cas nous avons $\tau^* = \tau_0$. En général, il n'y a pas de garantie que τ^* soit optimal dans la classe de densité instrumentale GED, *i.e.* que $\tau^* = \tau_0$.

1.3.2 Mesures de distorsion de risque

La VaR a été critiquée pour donner une vision trop limitée du niveau de risque réel. En particulier, la VaR ne dit rien sur ce qui se passe lorsque les pertes dépassent la VaR. En plus, la VaR ne satisfait pas la propriété de sous-additivité (voir Artzner et al. (1999), Wirch et Hardy (1999)). L'Expected shortfall (ES) est une mesure du risque alternative qui contourne ce problème en mesurant la perte moyenne dans le cas de pertes supérieures à la VaR. En effet, l'ES est une DRM qui vérifie la propriété de sous-additivité (voir Wang (2000)).

La DRM conditionnelle est définie par

$$\mathrm{DRM}_t = \int_0^1 \mathrm{VaR}_t(u)dG(u), \tag{1.18}$$

dès lors que l'intégrale existe, où G est une fonction de distribution cumulative (CDF) sur [0,1] que l'on appelle la fonction de distorsion. La DRM peut être interprétée comme une somme pondérée des VaR, où les pondérations sont les augmentations de la fonction de distorsion. L'ES est obtenue avec $G(u) = (u/\alpha)1_{[0,\alpha[}(u) + 1_{[\alpha,\infty[}(u)$. D'autres exemples de DRM sont le DRM à hasard proportionnel, obtenu avec $G(u) = u^r$, et le DRM exponentiel obtenu avec $G(u) = (1-e^{ru})/(1-e^r)$, $r > 0$. Supposons $\int_0^1 \xi_u dG(u) < 0$, sous **A1** nous avons

$$\mathrm{DRM}_t = -\sigma(\epsilon_{t-1},\epsilon_{t-2},\dots;\theta_0)\int_0^1 \xi_u dG(u) = \sigma(\epsilon_{t-1},\epsilon_{t-2},\dots;\theta_{0,G}),$$

où

$$\theta_{0,G} = H\left(\theta_0, -\int_0^1 \xi_u dG(u)\right)$$

peut être appelé le paramètre de risque de DRM conditionnelle. Un estimateur naturel de ce paramètre est

$$\hat{\theta}^*_{n,G} = H\left(\hat{\theta}^*_n, -\int_0^1 \hat{\xi}^*_{n,u} dG(u)\right).$$

Pour estimer la VaR conditionnelle, $-\sigma(\epsilon_{t-1},\epsilon_{t-2},\ldots;\theta_0)\xi_u$, la densité instrumentale optimale h ne dépend pas de u. Pour estimer la VaR pondérée, $\mathrm{DRM}_t = -\sigma(\epsilon_{t-1},\epsilon_{t-2},\ldots;\theta_0)\int_0^1 \xi_u dG(u)$, il est donc naturel de choisir la même densité instrumentale optimale h, qui minimise τ_h, au moins dans le cas APARCH (voir Théorème 1.2).

1.3.3 Illustrations numériques

Nous considérons d'abord un cadre théorique dans lequel la distribution de η_t est supposée connue. Considérons deux classes de densités instrumentales, la GED (κ) et la distribution Student St_ν. Nous avons déterminé les densités instrumentales optimales au sein de chaque classe. Nous avons également comparé ces dernières avec la densité gaussienne en terme d'efficacité asymptotique relative. Des expériences de Monte Carlo ont été utilisées pour comparer les performances des différentes procédures d'estimation de la VaR pour des échantillons finis. Finalement, des illustrations sur des séries financières ont été proposées. Ces illustrations numériques montrent que les résultats de la théorie asymptotique s'appliquent avec succès en échantillons finis, et aboutissent à des gains d'efficacité importants.

1.4 Résultats du chapitre 3

Dans ce chapitre, nous nous référons au travail de Francq et al. (2011) qui ont étudié l'estimation du modèle GARCH(p,q) (1.4) en utilisant un gQML fondé sur la densité gaussienne généralisée de paramètre r. Cette densité coïncide avec la densité gaussienne pour $r = 2$. Ces auteurs ont testé l'efficacité du QMLE gaussien par rapport à celle du gQMLE non gaussien. De même nous proposons de tester l'efficacité de l'estimateur du paramètre VaR. Nous appliquons le gQMLE proposé dans le chapitre précédent pour le cas du modèle (1.4). Ce qui nous distingue des travaux existants est que nous estimons le paramètre de VaR conditionnelle (qui inclut à la fois un paramètre de volatilité et un quantile) alors que Francq et al. (2011) considèrent l'estimation de la variance conditionnelle. De plus, nous utilisons comme classe de densités instrumentales la dgG(b,p,d). Cette classe de densités est plus large que celle proposée par Francq et al. (2011).

En appliquant le Lemme 1.1, Théorème 1.1 et Corollaire 1.1 du chapitre 2 au modèle GARCH(p,q) usuel (1.4) nous obtenons le résultat suivant.

Lemme 1.2 (Intervalle de confiance de la VaR) *Sous l'hypothèse* **C**, $\hat{\theta}^*_{n,\alpha}$ *converge presque sûrement vers* $\theta_{0,\alpha}$ *quand* $n \to \infty$ *et*

$$\sqrt{n}\left(\hat{\theta}^*_{n,\alpha} - \theta_{0,\alpha}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, G_h \Sigma^h G_h'\right).$$

En utilisant la méthode delta, nous obtenons un intervalle de confiance pour $VaR_t(\alpha)$ au niveau statistique $\underline{\alpha}$. Notons que le risque statistique $\underline{\alpha}$ est de nature différente du risque financier α. Le risque statistique est celui de mal estimer la VaR à cause d'un modèle erroné, d'une méthode d'estimation inadéquate ou encore d'un manque de données historiques suffisantes pour pouvoir estimer les paramètres avec précision. Le risque financier est lié au DGP, et ne peut être modifié par l'économètre (sauf en recherchant un ensemble d'information pour lequel le risque conditionnel serait moindre). Soit $\hat{\Xi}_\alpha$ un estimateur consistant de la variance asymptotique $G_h \Sigma^h G_h{}'$, et $\mu_{\underline{\alpha}}$ le quantile d'ordre $(1 - \underline{\alpha}/2)$ de la distribution normale. On obtient alors comme intervalle de confiance de $VaR_t(\alpha)$

$$S_n := \left[V\hat{a}R_t(\alpha) \pm \frac{\mu_{\underline{\alpha}}}{\sqrt{n}} \sqrt{\frac{\partial \hat{\sigma}_t(\hat{\theta}_{n,\alpha})}{\partial \theta'} \hat{\Xi}_\alpha \frac{\partial \hat{\sigma}_t(\hat{\theta}_{n,\alpha})}{\partial \theta}} \right]. \tag{1.19}$$

1.4.1 Gain d'efficacité du gQMLE basé sur la distribution Double Gamma Généralisée par rapport au QMLE gaussien

Dans le cadre de l'estimation de la VaR conditionnelle, nous voulons tester l'efficacité du gQMLE basé sur une densité instrumentale h donnée à un niveau de risque $\underline{\alpha}$. Notons que la distribution optimale h est celle qui minimise τ_h donné par (1.11). Ainsi nous obtenons le τ_h optimal noté τ_{hopt}, avec

$$\tau_{hopt} = \min_{h \in \mathbb{H}} \frac{4Eg_1^2(\sigma_*^{-1}\eta_0,1)}{\left\{Eg_2(\sigma_*^{-1}\eta_0,1)\right\}^2}. \tag{1.20}$$

Soit h la distribution Double Gamma généralisée $\Gamma(b,p,d)$, notée dgG, de paramètres $b > 0$, $p > 0$ et $d > 0$, définie par

$$h(x) = h_{dgG}(x) = \frac{db^p}{2\Gamma(\frac{p}{d})} |x|^{p-1} e^{-|bx|^d}.$$

La distribution dgG peut être considérée comme une classe importante de densités instrumentales $\mathbb{H}_{dgG}$. Elle contient, en particulier, la répartition de Laplace

$$h_l(x) = e^{-|x|}/2,$$

avec un paramètre de position nul et un paramètre d'échelle égal à 1, qui correspond à la densité $\Gamma(1,1,1)$. La distribution GED de paramètre $\kappa > 0$, notée par GED(κ), coïncide avec $\Gamma((\frac{1}{2})^{\frac{1}{\kappa}},1,\kappa)$ et elle est définie par

$$h_{GED_\kappa}(x) = \frac{\kappa}{\Gamma(1/\kappa) 2^{1+1/\kappa}} e^{-\frac{|x|^\kappa}{2}}.$$

Nous notons aussi que la distribution gaussienne est la distribution $\Gamma(\frac{1}{\sqrt{2}},1,2)$ ayant la forme

$$h(x) = \phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}.$$

La distribution double Weibull (Dweib) de paramètre de forme $k > 0$ et paramètre d'échelle $\lambda > 0$, définie par

$$h_{Dweib}(x) = \frac{1}{2}\frac{k}{\lambda}|\frac{x}{\lambda}|^{k-1}e^{-|\frac{x}{\lambda}|^k},$$

coïncide avec $\Gamma(\frac{1}{\lambda},k,k)$. En plus des distributions citées ci-dessus, la dgG contient également les distributions Rayleigh et Maxwell.

Pour $h \in \mathbb{H}_{dgG}$, et $x \neq 0$, nous avons

$$h'_{dgG}(x) = \frac{p-1-d|bx|^d}{|x|}h_{dgG}(x),$$

$$x\frac{h'_{dgG}}{h_{dgG}}(x) = p-1-d|bx|^d,$$

et

$$\sigma_* = \left(\frac{db^d}{p}E|\eta_1|^d\right)^{1/d}.$$

Nous obtenons alors,

$$\tau_{dgG} = \tau(d) = \frac{4}{d^2}\left(\frac{E\left|\eta_1\right|^{2d}}{\left(E\left|\eta_1\right|^d\right)^2} - 1\right). \tag{1.21}$$

Étant donnée l'équation (1.20), τ_{dgG} qui correspond à la densité instrumentale dgG optimale, notée τ^*_{dgG}, est définie par

$$\tau^*_{dgG} = \min_{h\in\mathbb{H}_{dgG}} \frac{4}{d^2}\left(\frac{E\left|\eta_1\right|^{2d}}{\left(E\left|\eta_1\right|^d\right)^2} - 1\right).$$

Soit pour tout $d > 0$, $\mu_d = E|\eta_t|^d$, alors

$$\tau^*_{dgG} = \min_{h\in\mathbb{H}_{dgG}} \frac{4}{d^2}\left(\frac{\mu_{2d}}{{\mu_d}^2} - 1\right).$$

Le problème d'optimisation ne dépend donc que du paramètre d et de certains moments du processus d'innovation. Si les moments théoriques sont remplacés par les moments empiriques et si l'espace des valeurs possibles pour d n'est pas borné, on peut montrer que la solution du problème d'optimisation diverge (comme dans Francq et al. (2011), lemme 3.1). Pour résoudre ce problème, nous devons faire l'hypothèse suivante.

B1: Il existe un unique $d_{opt} > 0$ tel que

$$d_{opt} = \arg\min_{d\in\mathbb{D}} \frac{4}{d^2}\left(\frac{E\left|\eta_1\right|^{2d}}{\left(E\left|\eta_1\right|^{d}\right)^2} - 1\right),$$

où $\mathbb{D}$ est un espace compact de la forme $\mathbb{D} = [\underline{d},\bar{d}] \in (0,d_{max})$ avec $d_{max} = \sup\left\{d \in \mathbb{R}; \mu_{2d} < \infty\right\}$.

Pour $u > 0$ et $v \geq 0$, on note

$$m(u,v) = E\left\{|\eta_t|^u(log|\eta_t|)^v\right\}, \tag{1.22}$$

lorsque les espérances existent.

B2: Le support Ω_η de η_t contient au moins cinq valeurs, et

$$E\left|\eta_t^{4d_{opt}}(log|\eta_t|)^2\right| < \infty.$$

Par convention nous poserons

$$0^u(log|0|)^v = 0. \tag{1.23}$$

Soit

$$m_n(u,v,\theta) = \frac{1}{n}\sum_{t=1}^{n}|\frac{\epsilon_t}{\sigma_t(\theta)}|^u(log|\frac{\epsilon_t}{\sigma_t(\theta)}|)^v. \tag{1.24}$$

et

$$\tau_n(d,\theta) = \frac{4}{d^2}\left(\frac{m_n(2d,0,\theta)}{m_n(d,0,\theta)^2} - 1\right). \tag{1.25}$$

En remplaçant dans (1.24), $\frac{\epsilon_t}{\sigma_t(\theta)}$ par la statistique $\frac{\epsilon_t}{\widetilde{\sigma}_t(\theta)}$ ou par le résidu $\hat{\eta}_t = \frac{\epsilon_t}{\widetilde{\sigma}_t(\hat{\theta}_{n,\phi})}$ du QMLE gaussien, nous obtenons respectivement

$$\widetilde{m}_n(u,v,\theta) = \frac{1}{n}\sum_{t=1}^{n}|\frac{\epsilon_t}{\widetilde{\sigma}_t(\theta)}|^u(log|\frac{\epsilon_t}{\widetilde{\sigma}_t(\theta)}|)^v$$

et

$$\hat{m}_n(u,v) = \frac{1}{n}\sum_{t=1}^{n}|\hat{\eta}_t|^u(log|\hat{\eta}_t|)^v.$$

Nous pouvons alors définir

$$\widetilde{\tau}_n(d,\theta) = \frac{4}{d^2}\left(\frac{\widetilde{m}_n(2d,0,\theta)}{\widetilde{m}_n(d,0,\theta)^2} - 1\right).$$

et, en prenant la limite quand $n \to \infty$ dans (1.24),

$$m_\infty(u,v,\theta) = E|\frac{\epsilon_t}{\sigma_t(\theta)}|^u(log|\frac{\epsilon_t}{\sigma_t(\theta)}|)^v.$$

L'hypothèse suivante sera utilisée pour obtenir un estimateur asymptotiquement normal de d_{opt}.

B3: d_{opt} appartient à l'intérieur de $\mathbb{D}$.

Soit $\hat{\eta}_t = \epsilon_t/\tilde{\sigma}_t(\hat{\theta}_n)$, $t = 1,\dots,n$, où $\hat{\theta}_n$ est le QMLE gaussien.

$$\hat{d}_n = \arg\min_{d\in\mathbb{D}} \frac{1}{d^2}\left(\frac{\hat{\mu}_{2d}}{\hat{\mu}_d^2} - 1\right), \quad \hat{\mu}_r = \frac{1}{n}\sum_{t=1}^{n}|\hat{\eta}_t|^r. \tag{1.26}$$

Pour tout n fixé,

$$\frac{1}{d^2}\left(\frac{\hat{\mu}_{2d}}{\hat{\mu}_d^2} - 1\right) \to 0, \quad \text{quand } d \to \infty.$$

C'est pour cette raison que nous supposons dans **B1** que d appartient à un intervalle borné $\mathbb{D}$. Le théorème suivant montre les propriétés de consistance et de normalité asymptotique de $\hat{d}_n$.

Théorème 1.3 *Sous les hypothèses **B1, B2, B3** et **C**, nous avons,*

$$\hat{d}_n \quad \to \quad d_{opt} \quad p.s.$$

et

$$\sqrt{n}(\hat{d}_n - d_{opt}) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Upsilon_{d_{opt}}). \tag{1.27}$$

Nous montrons que $\Upsilon_{d_{opt}}$ est de la forme

$$\Upsilon_{d_{opt}} = \frac{\zeta_{d_{opt}}}{\left\{\frac{\partial^2\tau_\infty}{\partial d^2}(d_{opt},\theta_0)\right\}^2}. \tag{1.28}$$

Le résultat suivant permet de tester si le QMLE gaussien est asymptotiquement plus précis que le gQMLE basé sur la densité instrumentale dgG.

Corollaire 1.3 *On se place sous les hypothèses du Théorème 1.3. Au niveau de confiance asymptotique $1-\underline{\alpha}$, l'hypothèse nulle que $d_{opt} = 2$ est rejetée si*

$$\left\{\frac{n}{\hat{\Upsilon}_{d_n}}(\hat{d}_n - 2)^2 > \left\{\Phi^{-1}(1-\underline{\alpha}/2)\right\}^2\right\}, \tag{1.29}$$

où $\hat{\Upsilon}_{d_n} = \frac{\hat{\zeta}_{d_{opt}}}{\left\{\frac{\partial^2\hat{\tau}_\infty}{\partial d^2}(d_{opt},\theta_0)\right\}^2}$ et Φ est la fonction de répartition de la loi normale standard.

Le corollaire ci-dessous donne un intervalle de confiance asymptotique pour le paramètre d_{opt}.

Corollaire 1.4 *Soit $\hat{d}_n$ un estimateur consistant du paramètre optimal inconnu d_{opt} et $\hat{\Upsilon}_{d_n}$ un estimateur consistant de la variance asymptotique $\Upsilon_{d_{opt}}$. L'ensemble*

$$I_n := \left[\hat{d}_n - \Phi^{-1}_{1-\underline{\alpha}/2} \sqrt{\frac{\hat{\Upsilon}_{d_n}}{n}} ; \hat{d}_n + \Phi^{-1}_{1-\underline{\alpha}/2} \sqrt{\frac{\hat{\Upsilon}_{d_n}}{n}} \right] \tag{1.30}$$

est un intervalle de confiance pour d_{opt} au niveau de risque asymptotique $\underline{\alpha} \in]0,1[$, c'est-à-dire

$$\lim_{n \to \infty} P[d_{opt} \in I_n] = 1 - \underline{\alpha}.$$

1.4.2 Résultats numériques

Nous commençons par présenter une application théorique dans laquelle la densité des innovations est supposée connue. Nous déterminons numériquement le paramètre τ_h intervenant dans la variance asymptotique de l'estimateur du paramètre de VaR (voir le théorème 1.2), en fonction de la densité instrumentale h choisie pour le gQMLE, et nous comparons cette valeur à celle obtenue pour le QMLE gaussien. Ceci illustre le fait que, lorsque la loi de η_t n'est pas gaussienne, un estimateur du paramètre de VaR fondé sur un gQMLE permet d'obtenir un gain d'efficacité asymptotique. Lorsque la loi de η_t est gaussienne, l'utilisation d'une classe de gQMLE qui ne contient pas la loi normale peut entraîner une perte d'efficacité asymptotique. Même si la classe de gQMLE contient la loi normale comme cas particulier (ce qui est vrai pour les densités instrumentales dgG) on peut s'attendre à une perte d'efficacité à distance finie, liée au fait que l'on doit utiliser (1.26) pour estimer le paramètre d_{opt} qui correspond à la densité instrumentale asymptotiquement optimale (et qui vaut 2 dans ce cas). Il est donc intéressant en pratique de tester si le gQMLE optimal est gaussien ou pas, c'est-à-dire si $d_{opt} = 2$ ou $d_{opt} \neq 2$.

Nous effectuons ensuite des simulations qui montrent que le test de l'hypothèse $d_{opt} = 2$ défini dans le corollaire 1.3 se comporte bien en échantillon fini. Nous appliquons également le Corollaire 1.4 pour donner des intervalles de confiance pour le paramètre d.

Enfin, nous appliquons le test proposé pour différentes séries de données financières telles que les indices boursiers, les taux de change et les matières premières.

Nous tirons les mêmes conclusions que celles tirées au niveau de simulations. En plus nous proposons un deuxième test, où on fixe $d = 1$. Nous trouvons pour les trois séries que cette hypothèse est acceptée. Nous suggérons donc que la densité instrumentale dgG de paramètre $d = 1$ peut être utilisée pour estimer la VaR conditionnelle de nombreuses séries financières.

1.5 Résultats du chapitre 4

Dans ce chapitre, nous étudions le concept de risque de modèle associé aux procédures paramétriques d'estimation de la VaR conditionnelle. Le risque de modèle peut affecter la volatilité et la VaR conditionnelle soit suite à des erreurs de spécification ou risques d'estimation. Par conséquent, les modèles non-paramétriques peuvent être considérés comme modèles alternatifs, étant donné qu'ils ne reposent pas sur une forme explicite de la volatilité. Nous considérons les modèle SVR basés sur la fonction de perte moindres carrés (LS-SVR), le modèle LS-SVR pondéré (WLS-SVR) et le modèle LS-SVR de taille fixe (FS-LS-SVR) afin de traiter le problème de l'estimation de la VaR conditionnelle en tenant compte du risque de modèle.

1.5.1 Risque de modèle lié à l'estimation paramétrique de la volatilité

L'objectif de cette partie est de détailler le concept du risque de modèle lié à l'estimation de la volatilité et de la VaR conditionnelle.

Par définition, le risque de modèle résulte du fait que le DGP est inconnu. Crouhy et al. (1998) ont indiqué que le risque de modèle est considéré comme le risque induit par la spécification et l'estimation des modèles statistiques. Conformément à la définition de Crouhy et al. (1998) la spécification des modèles économétriques pourrait être affectée par une erreur résultant soit de l'omission de certaines variables explicatives pertinentes ou d'une forme fonctionnelle inconnue.

En effet, les modèles de type ARCH/GARCH sont utilisés pour la modélisation de séries financières. De ce fait, il devient difficile d'identifier si la spécification est correcte ou si le modèle réel est disponible en tant qu'un des modèles candidats. Par conséquent, l'erreur de spécification du modèle de la volatilité ne peut être évitée car le DGP est inconnu, aussi la forme de la distribution des innovations n'est pas souvent gaussienne. Plus explicitement, les rendements des actifs n'ont pas suffisamment d'informations répondant à une question quel est le meilleur modèle de volatilité à utiliser.

Il est important de noter que les séries de données financières ont beaucoup de faits stylisés tels que l'asymétrie, la leptockurtocité, la distribution non-normale, la non-stationnarité ainsi de suite. Pour surmonter le problème susmentionné, Nelson (1991) a développé le modèle EGARCH et Glosten et al. (1993) ont introduit le modèle GJR. Ces modèles prennent en considération l'effet de levier. Une autre

façon de modéliser les asymétries est de considérer le modèle TGARCH introduit par Zakoïan (1994). Ce n'est pas facile de trouver un modèle adéquat en raison de l'existence des faits stylisés.
L'inférence des modèles ci-dessus est généralement basée sur la théorie du maximum de vraisemblance (ML). Cette approche convient pour le cas d'une distribution normale, mais elle n'est pas valide dans des applications réelles. Ainsi, le quasi maximum de vraisemblance (QML) a été utilisé pour estimer les paramètres des modèles GARCH. Le principe de QML est basé sur la maximisation d'une forme simplifiée de la fonction de log-vraisemblance d'un modèle mal spécifié. La spécification du modèle consiste à sélectionner une forme fonctionnelle pour capturer la structure de dépendance entre les prix passés et futurs du marché financier. Dans ce contexte, la technique statistique semi-paramétrique QML est adéquate (voir Francq et Zakoïan (2012)).
De même, les estimateurs non-paramétriques de la volatilité tels que l'approche SVR et ses variants sont conçus pour faire face à une forme fonctionnelle inconnue basée uniquement sur un ensemble d'apprentissage de N paires d'input-output (x_i,y_i). Par conséquent, elle résout le problème de l'erreur de spécification et permet d'éviter le risque d'estimation, c'est-à-dire, si l'hypothèse de normalité ou d'autres distributions spécifiques ne sont pas valides.

1.5.2 Estimation non-paramétrique de volatilité et de la VaR

Généralement les M-estimateurs vérifient la propriété de normalité asymptotique tels que le MLE et le QMLE. Hable (2012) a montré que le modèle SVR basé sur la fonction de perte LS peut être considéré comme un M-estimateur dont il vérifie la normalité asymptotique étant donné que la différence entre SVR empirique et SVR théorique est asymptotiquement normale. Ainsi nous nous basons sur cette propriété dans le choix de la fonction de perte.

Nous explicitons dans cette partie un détail de la littérature concernant les modèles LS-SVR, WLS-SVR et FS-LS-SVR.

Modèle LS-SVR

Soient X et Y deux variables aléatoires qui appartiennent à l'espace de probabilité (Ω,A,P). Soient χ et Υ un sous-ensemble fermé et borné de $\mathbb{R}^n$ avec une tribu $B(\chi)$ et un sous-ensemble fermé $\mathbb{R}$ avec un tribu$B(\Upsilon)$, respectivement. Soit l'ensemble finie $D_n = (x_1,y_1),...,(x_n,y_n)$ de taille n désigne une réalisation de variables aléatoires X et Y respectivement. Notons que (X_i,Y_j) pour $i = j$ sont iid avec une distribution inconnue F_{XY}. Étudier l'effet de X sur Y revient à déterminer la fonction de prévision optimale f. Le modèle LS-SVR définit la fonction f comme suit

$$\begin{cases} f : \chi \to \Upsilon, \\ f(x) = \omega^T \varphi(x) + b, \\ y_i = f(x_i) + e_i. \end{cases} \tag{1.31}$$

avec $\omega \in \mathbb{R}^n$, $\varphi(.)$ est une fonction de transformation non-linéaire, b est le processus de bruit et e_i sont les variables d'erreurs.

Au niveau de l'espace primal l'objectif est de déterminer les paramètres optimaux ω et b qui minimisent le risque empirique

$$R_{emp}(\omega,b) = \frac{1}{n} \sum_{i=1}^{n} (\omega^T \varphi(x_i) + b - y_i)^2, \tag{1.32}$$

sous la contrainte $\|\omega\|^2 \leq a$, et $a \in \mathbb{R}_+$.

D'après Suykens et al. (2002), le problème d'optimisation LS-SVR est défini comme suit

$$\begin{bmatrix} \underset{\omega,b,e}{min} J_p(\omega,e), J_p(\omega,e) = \frac{1}{2}\omega^T\omega + \gamma\frac{1}{2}\sum_{t=1}^{N} e_i^2 \\ \text{tel que} \quad y_i = \omega^T\varphi(x_i) + b + e_i, \quad i = 1,...,N \end{bmatrix}, \tag{1.33}$$

notons que $\varphi(.) : \mathbb{R}^n \to \mathbb{R}^{nh}$ est une cartographie non-linéaire dans l'espace à noyau, J est la fonction de perte, $\gamma \in \mathbb{R}^+$ est une constante réglable considérée comme un facteur de pénalité et $e_i \in \mathbb{R}$ sont les variables d'erreur. Pour résoudre le problème d'optimisation dans l'espace dual, nous définissons la fonction de Lagrange

$$L(\omega,b,e) = J_p(\omega,e) - \sum_{t=1}^{N} \alpha_i(\omega^T\varphi(x_i) + b + e_i - y_i), \tag{1.34}$$

où $\alpha_i \in \mathbb{R}$ sont les multiplicateurs de Lagrange connus comme vecteurs de support. Selon les conditions KKT, les conditions d'optimalité sont données par

$$\begin{cases} \dfrac{\partial L}{\partial \omega} = 0 \to \omega = \displaystyle\sum_{t=1}^{N} \alpha_i \varphi(x_i), \\ \dfrac{\partial L}{\partial b} = 0 \to \displaystyle\sum_{t=1}^{N} \alpha_i = 0, \\ \dfrac{\partial L}{\partial e_i} = 0 \to \alpha_i = \gamma e_i, \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \to \omega^T\varphi(x_i) + b + e_i - y_i = 0, \quad i = 1, \ldots, N \end{cases} \tag{1.35}$$

nous avons,

$$\begin{bmatrix} \text{résoudre dans} \quad \alpha,b: \\ \left[\frac{0}{1_v}\Big|\frac{1_v^T}{\Omega+I/\gamma}\right]\left[\frac{b}{\alpha}\right]=\left[\frac{0}{y}\right] \end{bmatrix}, \tag{1.36}$$

avec, $y=[y_1,...,y_N]$, $I_v=[1,...,1]$, $\alpha=[\alpha_1,...,\alpha_N]$ et Ω est la matrice Hessiene $N\times N$.
Nous avons

$$\omega_{ij}=\varphi^T(x_i)\varphi(x_j)=K(x_i,x_j), \qquad i,j=1,\ldots,N.$$

Selon le théorème de Mercer, le modèle LS-SVR devient

$$f(x)=\sum_{i=1}^{N}\alpha_i K(x,x_i)+b. \tag{1.37}$$

En pratique, la fonction du noyau la plus utilisée est la fonction RBF donnée par

$$K(x,x_i)=exp\left(-\left\|x-x_i\right\|^2/2\sigma^2\right). \tag{1.38}$$

Modèle WLS-SVR

Selon à Suykens et al. (2002) et Hou et al. (2013), dans l'espace $\mathbb{R}^n$ l'équation d'optimisation du modèle WLS-SVR est donnée par

$$\begin{bmatrix} \underset{\omega,b,e}{min} J_p(\omega,e), J_p(\omega,e)=\frac{1}{2}\omega^T\omega+\gamma\frac{1}{2}\sum_{t=1}^{N}\nu_i e_i^2 \\ \text{tel que} \quad y_i=\omega^T\varphi(x_i)+b+e_i. \quad i=1,...,N \end{bmatrix}. \tag{1.39}$$

Pour résoudre le système d'équations ci-dessus, la fonction de Lagrange peut être écrite comme suit

$$L(\omega,b,e)=J_p(\omega,e)-\sum_{t=1}^{N}\alpha_i(\omega^T\varphi(x_i)+b+e_i-y_i).$$

Selon les conditions d'optimisation KKT, les conditions d'optimalité dans l'équation (1.39) sont données par

$$
\begin{cases}
\dfrac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{t=1}^{N} \alpha_i \varphi(x_i), \\
\dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{t=1}^{N} \alpha_i = 0, \\
\dfrac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \dfrac{\gamma \nu_i e_i}{N}, \\
\dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow \omega^T \varphi(x_i) + b + e_i - y_i = 0, \qquad i = 1,\ldots,N.
\end{cases}
\tag{1.40}
$$

En éliminant la constante ω et e_i le système d'équations (1.40) s'écrit comme suit

$$
\begin{bmatrix} \text{résoudre dans} \quad \alpha,b : \\ \left[\frac{0}{1_v} \middle| \frac{1_v^T}{\Omega + V_\gamma}\right] \left[\frac{b}{\alpha}\right] = \left[\frac{0}{y}\right] \end{bmatrix},
\tag{1.41}
$$

où la matrice diagonale V_γ est donnée par $V_\gamma = diag\left(\left[\frac{1}{\gamma v_1},...,\frac{1}{\gamma v_N}\right]\right)$, $\alpha = [\alpha_1,...,\alpha_N]^T$ et $y = [y_1,...,y_N]^T$. En résolvant le système d'équations (1.40), nous obtenons les paramètres α_i. Le modèle WLS-SVR s'écrit

$$
f(x) = \sum_{i=1}^{N} \alpha_i K(x,x_i) + b.
\tag{1.42}
$$

Suykens et al. (2002), on introduit le facteur de pondération v_i tel que

$$
\nu_i = \begin{cases} 1 & \text{si} \quad |\frac{e_i}{\hat{s}}| < c_1 \\ \frac{c_2 - |\frac{e_i}{\hat{s}}|}{c_2 - c_1} & \text{si} \quad c_1 < |\frac{e_i}{\hat{s}}| < c_2 \\ 10^{-4} & \text{sinon.} \end{cases}
\tag{1.43}
$$

Notons que $\hat{s}$ est donné par (Andrzej et Shunichi, 2006)

$$
\hat{s} = 1.483 Med \left\{ |e_i - Med|e_i|| \right\},
\tag{1.44}
$$

avec $c_1 = 2.5$ et $c_2 = 3$.

Modèle FS-LS-SVR

La méthode FS-LS-SVR résoud un système de M équations linéaires sur la base de l'approximation du Nyström en introduisant la parcimonie dans le modèle LS-SVR. Le principal problème dans l'espace primal est de donner une forme explicite de la fonction non-linéaire φ basée sur la décomposition de la fonction de noyau $K(x,x_i)$. Étant donné un ensemble de données $\{x_i,y_i\}_{i=1}^N$, les valeurs propres λ_i, les fonctions propres ϕ_i, leurs approximations $\hat{\lambda}_i^{(s)}$ et les vecteurs propres U_i alors la décomposition de la matrice de noyau Ω avec des entrés $K(x,x_i)$ devient possible.

Cette décomposition permet d'obtenir une approximation $\hat{\varphi}(x)$ pour tout i qui sert à estimer ω et b comme suit

$$\hat{\varphi}_i(x^{(\nu)}) = \frac{N}{\sqrt{\lambda_i^{(s)}}} \sum_{k=1}^{N} u_{ik} K(x_k,x^{(\nu)}). \tag{1.45}$$

L'approximation $\hat{\varphi}(x)$ à dimension finie peut être obtenue sur la base des critères quadratiques Renyi entropie avec une sélection de $M << N$ vecteurs de support. Ceci conduit à une représentation parcimonieuse dans l'espace primal.
L'algorithme pour la mise en oeuvre finale peut être décrit par les étapes suivantes proposées par Espinoza et al. (2004)

Etape 1: Normaliser les inputs et outputs pour avoir une moyenne nulle et une variance égale à l'unité.

Etape 2: Sélectionner un sous-échantillon initial de taille M.

Etape 3: Construire la matrice du noyau de taille M et calculer sa décomposition propre.

Etape 4: Construire une approximation de cartographie non-linéaire pour le reste de données.

Etape 5: Estimer une régression linéaire dans l'espace primal.

Etape 6: Estimer l'approximation de cartographie non-linéaire pour les données de validation.

Etape 7: Utiliser les estimations de régression avec la cartographie non linéaire des données de validation pour produire de la prévision.

Méthode proposée pour estimer la VaR

Dans ce chapitre nous proposons d'utiliser les modéles LS-SVR, WLS-SVR et FS-LS-SVR pour estimer la VaR. Le modéle LS-SVR peut être appliqué comme une fonction non-linéaire F de la série de rendements ϵ_t comme suit

$$\sigma_t^2 = F[\epsilon_{t-1},\epsilon_{t-2},...,\epsilon_{t-p}]. \tag{1.46}$$

Étant donnée la série de rendements ϵ_t définie par

$$\begin{cases} \epsilon_t = 100log(\frac{P_t}{P_{t-1}}) \\ \epsilon_t = \sigma_t \eta_t, \end{cases} \tag{1.47}$$

où P_t est le prix de fermeture à l'instant t, (η_t) est la séquence des innovations et σ_t est l'écart type. La forme log linéaire des rendements (1.47) est donnée par

$$\begin{cases} log\epsilon_t^2 = log\sigma_t^2 + (log\eta_t^2 - \mu) + \mu \\ log\epsilon_t^2 = f(\epsilon_{t-1}, ..., \epsilon_{t-p}) + z_t, \end{cases} \tag{1.48}$$

où, z_t est la séquence des innovations iid. Par conséquent, la volatilité estimée est donnée par

$$\hat{\sigma}_t^2 = e^{\hat{f}(\epsilon_{t-1}, ..., \epsilon_{t-p})}. \tag{1.49}$$

Suivant Gavrishchaka et Banerjee (2006), nous appliquons le modèle LS-SVR et ses variants au $log(\epsilon_t^2)$ au lieu de ϵ_t^2 suivie d'une application exponentielle aux outputs du modèle LS-SVR. Dans la pratique, les modèles LS-SVR, WLS-SVR et FS-LS-SVR nécessitent une spécification de volatilité dans l'équation (1.46) dont nous supposons que $\sigma_t^2 = \epsilon_{t-1}^2$.

En utilisant l'output de l'équation (1.49) et pour un niveau de risque $\alpha \in (0,1)$, nous aurons la forme de la VaR conditionnelle suivante

$$P_{t-1}\left[\epsilon_t < -VaR_t(\alpha)\right] = \alpha, \tag{1.50}$$

où P_{t-1} indique la distribution historique conditionnelle à $\{\epsilon_\mu, \mu < t\}$ lorsque (ϵ_t) satisfait l'équation (1.47). La VaR théorique est définie par

$$VaR_t(\alpha) = -\sigma_t(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-p}) F_z^{-1}(\alpha), \tag{1.51}$$

où $F_z^{-1}(\alpha)$ est la fonction de densité de z_t, $F_z^{-1}(\alpha) < 0$ pour des petites valeurs de α.

1.5.3 Applications

Au niveau de ce chapitre nous nous intéressons à souligner le risque de modèle lié à l'estimation de la VaR. En effet, l'objectif des simulations est d'étudier le risque lié aux modèles paramétriques. Tout d'abord nous simulons trois modèles GARCH avec différentes distributions des innovations. Ensuite, nous évaluons la qualité d'ajustement des modèles GARCH en utilisant la fonction de log vraisemblance (LLF) et le critère AIC. Nous utilisons aussi le test de rapport de vraisemblance ou le test de mauvaise spécification noté (LRtest). Ce test nous indique que si on accepte l'hypothèse nulle alors le modèle restreint est meilleur.

Suite aux résultats de simulations, nous avons pensé à appliquer seulement le modèle LS-SVR et ses variants pour estimer la VaR. Nous avons effectué un test de corrélation des rendements, où on a pu montrer qu'il s'agit des corrélations négatives qui indiquent dans le cadre d'analyse paramétrique il est nécessaire d'utiliser un modèle GARCH asymétrique pour tenir compte de cette notion. Mais comme nous le savons, on peut appliquer différents modèles asymétriques sans connaitre le meilleur d'entre eux.

Afin de vérifier les performances de prévision des modèles utilisés nous considérons les mesures de qualité d'ajustement tel que le MSE, RMSE et MAE. Le meilleur estimateur est celui qui a les valeurs les plus faibles de ces critères.

- $MSE = \frac{1}{N}\sum_{t=1}^{N}(\epsilon_t^2 - \hat{\sigma}_t^2)^2.$

- $RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\epsilon_t^2 - \hat{\sigma}_t^2)^2}.$

- $MAE = \frac{1}{N}\sum_{t=1}^{N}|\epsilon_t^2 - \hat{\sigma}_t^2|.$

Comme première étape, nous estimons la volatilité, puis nous estimons la VaR. Nous effectuons en dernier lieu une comparaison entre les modèles non-paramétriques utilisés suite au nombre des violations de la VaR par rapport à la VaR moyenne.

1.6 Bibliographie

Andrzej, C. and A. Shunichi (2005) Adaptive blind signal and image processing. *Publish. H. of electron. Indust.*

Artzner, P., Delbaen, F., Eber, J-M. and D. Heath (1999) Coherent measures of risk. *Mathematical Finance* 9, 203–228.

Bardet, J-M. and O. Wintenberger (2009) Asymptotic normality of the Quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics* 37, 2730–2759.

Berkes, I. and L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* 32, 633–655.

Berkes, I., Horváth, L. and P. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Candelon, B., Colletaz, G., Hurlin, C. and S. Topkpavi (2011) Backtesting Value at?Risk: A GMM Duration?Based Test. *Journal of Financial Econometrics*, 1–30.

Christoffersen, P. F. and D. Pelletier (2004) Backtesting Value-at-Risk: a duration-based approach. *Journal of Financial Econometrics* 2, 84–108.

Crouhy, M. Galai, D., and R. Mark (1998) Model risk. *Journal of Financial Engineering* 7, 267–288.

Ding, Z., Granger C. and R.F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

Dowd, K. (1998) *Beyond Value at Risk: The New Science of Risk Management.* Chichester and New York: John Wiley and Sons.

Dowd, K. and D. Blake (2006) After Var: the theory, estimation and insurance applications of quantile-based risk measures. *J. Risk Insur* 73(2), 193–229.

Dumitrescu, E-I., Hurlin, C. and V. Pham (2012) Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. Working Papers halshs–00671658, HAL.

Engle, R-F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987–1008.

Engle, R-F. and S. Manganelli (2004) CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics* 22, 367–381.

Espinoza, M., Pelckmans, K., Hoegaerts, L., Suykens, J.A.K., and B. De Moor (2004) A comparative study of LS-SVMs applied to the silver box identification problem. *Proc. Of the 6th IFAC Symposium on Nonlinear Control Systems (NOLCOS).*

Fisher, R-A. and L-H-C. Tippett (1928) Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Camb. Phil. Soc* 24, 180–190.

Francq, C., Lepage, G. and J-M. Zakoïan (2011) Two-stage non Gaussian QML estimation of GARCH Models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics* 165, 246–257.

Francq, C. and J-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. and J-M. Zakoïan (2012) Risk-parameter estimation in volatility models. MPRA Preprint No. 41713.

Francq, C. and J-M. Zakoïan (2013) Optimal predictions of powers of conditionally heteroskedastic processes. *Journal of the Royal Statistical Society - Series B* 75, 345–367.

Gavrishchaka, V-V. and S. Banerjee (2006) Support vector machine as an efficient framework for stock market volatility forecasting. *Comput Manag Sci* 3(2), 147–160.

Glosten, L., Jagannathan, R. and D. Runkle (1993) Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

Gyorfi, L., Kohler, M., Krzyzak, A. and H. Walk (2002) *A Distribution-Free Theory of Nonparametric Regression*, New York: Springer.

Hable R. (2012) Asymptotic Normality of Support Vector Machine Variants and Other Regularized Kernel Methods. *Journal of Multivariate Analysis* 106, 92–117.

Hamadeh, T. and J-M. Zakoïan (2011)Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH Processes. *Journal of Statistical Planning and Inference* 141, 488–507.

Hou, L., Yang, S., Wang, X. and J. Shen (2013) Short Term Load Forecasting Based on WLS-SVR and TGARCH Error Correction Model in Smart Grid. *PRZEGLD ELEKTROTECHNICZNY, ISSN 0033-2097, R. 89 NR 3b.*

Jorion, P. (2001) *Value-at-risk: the new benchmark for managing financial risk.* McGraw-Hill, New York, 2nd edition.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3, 73–84.

Mikosch, T. and D. Straumann (2006) Stable limits of martingale transforms with application to the estimation of GARCH parameters. *The Annals of Statistics* 34, 493–522.

Nelson, D-B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59, 347–370.

Steinwart, I. and A. Christmann (2008) *Support vector machines*, Springer, New York.

Suykens, J-A-K., Brabanter, J-D. and L. Lukas (2002) Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing* 48, 85–105.

Vapnik, V-N. (1998) *Statistical Learning Theory.* John Wiley and Sons, New York.

Wang, S. (2000) A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance* 67, 15–36.

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

Wirch, J.L. and M.R. Hardy (1999) A Synthesis of Risk Measures for Capital Adequacy. *Insurance: Mathematics and Economics* 25, 337–347.

Zakoïan, J-M. (1994) Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control* 18, 931–955.

Chapitre 2

Consistent estimation of the Value at Risk when the error distribution of the volatility model is misspecified

Abstract. A two-step approach for conditional Value at Risk (VaR) estimation is considered. First, a generalized-quasi-maximum likelihood estimator (gQMLE) is used to estimate the volatility parameter, then the empirical quantile of the residuals serves to estimate the theoretical quantile of the innovations. When the instrumental density h of the gQMLE is not the Gaussian density, or the true distribution of the innovations, then both the estimations of the volatility and of the quantile are asymptotically biased.. The two errors however counterbalance each other, and we finally obtain a consistent estimator of the conditional VaR. For different GARCH models, we derive the asymptotic distribution of the VaR estimation based on gQMLE. We show that the optimal instrumental density h depends neither on the GARCH parameter nor on the risk level, but only on the distribution of the innovations. A simple adaptive method based on empirical moments of the residuals allows to infer an optimal element within a class of potential instrumental densities. Important asymptotic efficiency gains are achieved by using gQMLE when the innovations are heavy-tailed. We extended our approach to Distortion Risk Measure parameter estimation, where consistency of the gQMLE is proved. Simulation experiments and real case study are provided.

KEYWORDS. APARCH, Conditional VaR, Distortion Risk Measures, GARCH, Generalized Quasi Maximum Likelihood Estimation, Instrumental density.

NOTE. The content of this chapter is based on a paper written in collaboration with Christian Francq and Mohamed EL Ghourabi.

2.1 Introduction

Financial market risk is usually perceived as the exposure to potential losses of portfolios of risky assets. To assess the risk level, practitioners rely on risk management tools, such as the notorious Value-at-Risk (VaR). In the late 1980, financial firms began the use of VaR, defined as the loss that should not be reached for a given position over a holding time period and at a certain confidence level.

The VaR is often estimated by a simple quantile of the historical returns. This practice implicitly assumes that the sequence of the returns is stationary, and neglects the dynamics, in particular this does not account for the existence of clusters of extreme returns. It is preferable to take into account the information available, by reasoning on the conditional distribution of the returns (see *e.g.* McNeil, Frey and Embrechts (2005), Kuester, Mittnik and Paolella (2006) and Escanciano and Olmo (2010), who clearly showed that unconditional models of VaR are outperformed by conditional ones). The VaR conditional on past observations will be called the conditional VaR[1].

More precisely, at the risk level $\alpha \in (0,1)$, the (conditional) VaR of a sequence of returns (ϵ_t) is the opposite of the α-quantile of the conditional distribution:

$$\mathrm{VaR}_t(\alpha) = -\inf\left\{x : P(\epsilon_{t+1} \leq x \mid \epsilon_u, u \leq t) \geq \alpha\right\}. \tag{2.1}$$

Assume that the returns follow the general conditionally heteroscedastic model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t = \sigma_t(\theta_0) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta_0) \end{cases} \tag{2.2}$$

where (η_t) is a sequence of independent and identically distributed (iid) random variables, η_t is independent of $\{\epsilon_u, u < t\}$, $\theta_0 \in \mathbb{R}^m$ is a parameter belonging to a compact parameter space Θ, and $\sigma : \mathbb{R}^\infty \times \Theta \to (0,\infty)$. The variable σ_t^2 is generally referred to as the volatility of ϵ_t. For this GARCH-type volatility model, we have

$$\mathrm{VaR}_t(\alpha) = -\sigma_t(\theta_0)\xi_\alpha, \tag{2.3}$$

where ξ_α is the α-quantile of the distribution P_η of the innovations. Note that model (2.2) is not identifiable without a scaling assumption on P_η. The standard identifiability assumption is $E\eta_t^2 = 1$, but we do not need to make this assumption in the present chapter.

A simple and widely used example of (2.2) is the GARCH(p,q) model of Engle (1982) and Bollerslev (1986), defined by

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega_0 + \sum_{i=1}^q \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j}\sigma_{t-j}^2 \end{cases} \tag{2.4}$$

[1]Sometimes the conditional VaR refers to another risk measure called the expected shortfall.

where $\omega_0 > 0, \alpha_{0i} \geq 0$, $\beta_{0j} \geq 0$. For the GARCH(1,1) model, we have $\sigma_t^2 = \sum_{i=1}^{\infty} \beta_{01}^{i-1}(\omega_0 + \alpha_{01}\epsilon_{t-i}^2)$, provided $\beta_{01} < 1$.

The most widely used estimator of ARCH-type models is arguably the Gaussian QMLE. The consistency and asymptotic normality (CAN) of this estimator requires only few regularity assumptions, and the standard identifiability condition $E\eta_t^2 = 1$ (see Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004) for the case of standard GARCH and ARMA-GARCH models, Mikosch and Straumann (2006), Straumann and Mikosch (2006), Bardet and Wintenberger (2009) for more general models). In the framework of standard GARCH models, Berkes and Horváth (2004) introduced generalized non-Gaussian QMLE (gQMLE) and established their CAN under alternative identifiability conditions. For the general model (2.2), Francq and Zakoïan (2013) (hereafter FZ) showed that particular gQMLE's lead to convenient one-step predictions of the powers $|\epsilon_t|^r$, $r \in \mathbb{R}$. Francq *et al.* (2011) constructed a two-step procedure based on a particular class of gQMLE for estimating standard GARCH(p, q) models. Independently, Fan *et al.* (2014) proposed, also for the standard GARCH estimation problem, a three-step quasi maximum likelihood procedure allowing for the use of a vast class of non-Gaussian likelihood functions (see also the discussions that follow this paper). Francq and Zakoïan (2012) propose a gQMLE which allows for estimating a conditional VaR in one step and compare this method with the more standard two-step method which consists in estimating the volatility parameter by Gaussian QMLE and the quantile of the innovations by the empirical quantile of the residuals.

In the present chapter, we extend the conditional VaR two-step evaluation method based on the Gaussian QMLE by investigating the use of gQMLE's based on a generic instrumental density h. It is well known that the standard Gaussian QMLE, which is based on the instrumental density $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, converges to the volatility parameter θ_0, under mild regularity conditions. Moreover the empirical α-quantile of the Gaussian QMLE residuals converges to ξ_α. Section 2.1.1 shows that, in a very general framework, the gQMLE converges to some parameter θ_0^*, which depends on h, P_η and θ_0. When $h \neq \phi$ or $h \neq f_\eta$, where f_η is the Lebesgue density associated to P_η, generally we have $\theta_0^* \neq \theta_0$, and the empirical α-quantile of the gQMLE residuals converges to $\xi_\alpha^* \neq \xi_\alpha$. The conditional VaR two-step estimator is however consistent because $\sigma_t(\theta_0)\xi_\alpha = \sigma_t(\theta_0^*)\xi_\alpha^*$. Section 2.1.2 studies the asymptotic distribution of this estimator, for the general model (2.2). Section 2.2 makes explicit the asymptotic distributions for an extension of the GARCH model (2.4). It is shown that the optimal instrumental density, *i.e.* the function h which minimizes the asymptotic variance of the VaR estimator, depends neither on the GARCH parameter θ_0 nor on the risk level α, but only on simple characteristics of P_η. It follows that a simple adaptive method based on empirical moments of the residuals makes it possible to infer which h

is optimal. The possibility of extending the estimation method to conditional Distortion Risk Measures (DRM) is discussed. The numerical illustrations are displayed in Section 2.3. Section 2.4 provides a conclusion. The proofs are collected in the Appendix. For the standard volatility models, the following assumption is satisfied.

A1: There exists a continuous function H such that for any $\theta \in \Theta$, for any $K > 0$, and any sequence $(x_i)_i$

$$K\sigma(x_1, x_2, \dots; \theta) = \sigma(x_1, x_2, \dots; H(\theta, K)).$$

In the GARCH(1,1) case, we have

$$K\sigma_t(\theta) = \sqrt{K^2\omega + K^2\alpha + \beta\sigma_{t-1}^2} = \sigma_t \{H(\theta_0, K)\}$$

where $H(\theta_0, K) = (K^2\omega_0, K^2\alpha_{01}, \beta_{01})'$. Assumption **A1** means that the parametric form of the volatility is stable by scaling, which is a highly desirable property for an ARCH model.

In view of (2.3) and **A1**, when $\xi_\alpha < 0$ we have

$$\mathrm{VaR}_t(\alpha) = -\sigma_{t+1}(\theta_0)\xi_\alpha = \sigma_{t+1}(\theta_{0,\alpha})$$

where $\theta_{0,\alpha} = H(\theta_0, -\xi_\alpha)$. The parameter $\theta_{0,\alpha}$ is called the VaR parameter in Francq and Zakoïan (2012).

In the next section, we show that the gQMLE generally converges to a parameter θ_0^* such that $\sigma_t(\theta_0^*) = \sigma^*\sigma_t(\theta_0)$, where $\sigma^* > 0$ depends on h and P_η. Thus the residuals of the gQMLE are approximations of η_t/σ^*. Consequently, the gQMLE of the volatility converges to $\sigma^*\sigma_t(\theta_0)$ and the empirical quantile of the gQMLE residuals converges to $\xi_\alpha^* = \xi_\alpha/\sigma^*$. Hence, the gQMLE of the VaR gives a consistent estimator of $\mathrm{VaR}_t(\alpha) = -\sigma_{t+1}(\theta_0^*)\xi_\alpha^*$.

2.1.1 Estimating the volatility parameter

Given observations $\epsilon_1, \dots, \epsilon_n$ of (2.2), and arbitrary initial values $\widetilde{\epsilon}_i$ for $i \leq 0$, let

$$\widetilde{\sigma}_t(\theta) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \widetilde{\epsilon}_0, \widetilde{\epsilon}_{-1}, \dots; \theta).$$

This random variable can be seen as a proxy of

$$\sigma_t(\theta) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \dots; \theta).$$

Given an *instrumental* density $h > 0$, consider the QML criterion

$$\widetilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(\epsilon_t, \widetilde{\sigma}_t(\theta)), \qquad g(x, \sigma) = \log \frac{1}{\sigma} h\left(\frac{x}{\sigma}\right), \tag{2.5}$$

and the (generalized) QMLE

$$\hat{\theta}_n^* = \arg\max_{\theta \in \Theta} \widetilde{Q}_n(\theta).$$

Throughout the text, starred symbols are used to designate quantities which depend on the instrumental density h. This estimator is the standard Gaussian QMLE if h is the standard Gaussian density ϕ. It is important to note that the parametric form of $\sigma_t(\cdot)$ is assumed to be correctly specified, but we do not make precise assumptions on the distribution of η_t. In particular, we do not assume that $\mathrm{Var}(\eta_t) = 1$. Consequently σ_t^2 only corresponds to the conditional variance $\mathrm{Var}(\epsilon_t \mid \epsilon_u, u < t)$ up to the unknown constant $\mathrm{Var}(\eta_t)$. To establish the CAN of $\hat{\theta}_n^*$, we make the following assumptions.

A2: (ϵ_t) is a strictly stationary and ergodic solution of (2.2), and there exists $s > 0$ such that $E|\epsilon_1|^s < \infty$.

A3: For some $\underline{\omega} > 0$, almost surely $\widetilde{\sigma}_t(\theta) \in (\underline{\omega}, \infty]$ for any $\theta \in \Theta$ and any $t \geq 1$. For $\theta_1, \theta_2 \in \Theta$, we have $\sigma_t(\theta_1) = \sigma_t(\theta_2)$ $a.s.$ if and only if $\theta_1 = \theta_2$.

Note that by **A2**

$$g(\epsilon_t, \sigma_t(\theta)) = g\left(\eta_t, \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}\right) - \log \sigma_t(\theta_0). \tag{2.6}$$

A4: The function $\sigma \to Eg(\eta_0, \sigma)$ takes its values in $[-\infty, +\infty)$ and has a unique maximum at some point $\sigma_* \in (0, \infty)$.

A5: The instrumental density h is continuous on $\mathbb{R}$, it is also differentiable, except possibly in 0, and there exist constants $\delta \geq 0$ and $C_0 > 0$ such that, for all $u \in \mathbb{R} \setminus \{0\}$, $|uh'(u)/h(u)| \leq C_0(1 + |u|^\delta)$ and $E|\eta_0|^{2\delta} < \infty$.

A6: There exist a random variable C_1 measurable with respect to $\{\epsilon_u, u < 0\}$ and a constant $\rho \in (0, 1)$ such that $\sup_{\theta \in \Theta} |\sigma_t(\theta) - \widetilde{\sigma}_t(\theta)| \leq C_1 \rho^t$ a.s.

Under **A1** and **A4**, define the parameter

$$\theta_0^* = H(\theta_0, \sigma_*). \tag{2.7}$$

A7: The parameter θ_0^* belongs to the compact parameter space Θ.

A8: The parameter θ_0^* belongs to the interior $\overset{\circ}{\Theta}$ of Θ.

A9: There exists no non-zero $x \in \mathbb{R}^m$ such that $x' \frac{\partial \sigma_t(\theta_0^*)}{\partial \theta} = 0, \quad a.s.$

A10: The function $\theta \mapsto \sigma(x_1, x_2, \dots; \theta)$ has continuous second-order derivatives, and

$$\sup_{\theta\in\Theta}\left\|\frac{\partial\sigma_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{\sigma}_t(\theta)}{\partial\theta}\right\|+\left\|\frac{\partial^2\sigma_t(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2\widetilde{\sigma}_t(\theta)}{\partial\theta\partial\theta'}\right\|\leq C_1\rho^t,$$

where C_1 and ρ are as in **A6**.

A11: h is twice continuously differentiable, except possibly at 0, with $|u^2\,(h'(u)/h(u))'\,| \leq C_0(1+|u|^\delta)$ for all $u \in \mathbb{R}\setminus\{0\}$ and $E|\eta_0|^\delta < \infty$, where C_0 and δ are as in **A5**.

A12: There exists a neighborhood $V(\theta_0^*)$ of θ_0^* such that

$$\sup_{\theta\in V(\theta_0^*)}\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta}\right\|^4,\quad \sup_{\theta\in V(\theta_0^*)}\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial^2\sigma_t(\theta)}{\partial\theta\partial\theta'}\right\|^2,\quad \sup_{\theta\in V(\theta_0^*)}\left|\frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)}\right|^{2\delta}$$

have finite expectations.

Most of these assumptions are similar to those of Berkes and Horváth (2004) and FZ.

Remark 2.1 *Note that* **A4** *is much less restrictive than the analog assumption in FZ, which requires a maximum at $\sigma_* = 1$ (see* **A3** *in FZ). Note also that we do not need any identifiability condition on η_t (such that $E\eta_t^2 = 1$). We need weaker assumptions because, in our framework, it will only be necessary to define the volatility up to an unknown multiplicative constant. Actually,* **A4** *is the same as Assumption 2 made by Fan* et al. *(2014) for their three-step estimation procedure. Our framework is however quite different because we consider the VaR parameter estimation whereas Fan* et al. *consider the GARCH parameter estimation.*

Remark 2.2 *In view of (2.24) below, under* **A5** *the parameter σ_* defined in* **A4** *is such that*

$$E\left\{\frac{\eta_0}{\sigma_*}\frac{h'}{h}\left(\frac{\eta_0}{\sigma_*}\right)\right\} = -1. \tag{2.8}$$

For the standard GARCH case, several assumptions can be made more explicit. The true value of the parameter is $\theta_0 = (\omega_0, \alpha_{01}, \dots, \beta_{0p})'$ and the generic element of Θ is denoted by $\theta = (\omega, \alpha_1, \dots, \beta_p)'$. It is well-known that a necessary and sufficient condition for the existence of a strictly stationary solution to (2.4) is $\gamma < 0$, where γ denotes the top-Lyapunov exponent of the model (see e.g. Francq and Zakoïan (2004)). Write $\gamma = \gamma(\theta_0)$ to emphasize that γ depends on θ_0 (and also on the law of η_1). Let $\mathcal{A}_\theta(z) = \sum_{i=1}^q \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^p \beta_j z^j$. In that framework the assumptions **A2**, **A3**, **A6**, **A9**, **A10** and **A12** are equivalent to:

C: $\gamma(\theta_0) < 0$; $\forall\theta \in \Theta$, $\sum_{j=1}^p \beta_j < 1$ and $\omega > \underline{\omega}$ for some $\underline{\omega} > 0$; $|\eta_0|$ has a non degenerate distribution; if $p > 0$, $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ have no common root, $\mathcal{A}_{\theta_0}(1) \neq 0$, and $\alpha_{0q} + \beta_{0p} \neq 0$.

The following lemma extends results obtained by Berkes and Horváth (2004) in the standard GARCH case, and by FZ under stronger assumptions (see Remark 2.1).

Lemma 2.1 (Asymptotic behavior of generalized QMLE) *If* **A1-A7** *are satisfied, then*

$$\hat{\theta}_n^* \to \theta_0^*, \quad a.s.$$

where θ_0^* is defined by (2.7). If, in addition, **A8-A12** *are satisfied and $Eg_2(\eta_0, 1) \neq 0$ then*

$$\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau_h J_*^{-1})$$

where

$$J_* = 4ED_t(\theta_0^*)D_t'(\theta_0^*) \quad \textit{and} \quad \tau_h = \frac{4Eg_1^2(\sigma_*^{-1}\eta_0, 1)}{\left\{Eg_2(\sigma_*^{-1}\eta_0, 1)\right\}^2}, \tag{2.9}$$

in which

$$D_t(\theta) = \frac{1}{\sigma_t(\theta)}\frac{\partial \sigma_t(\theta)}{\partial \theta}, \quad g_1(x,\sigma) = \frac{\partial g(x,\sigma)}{\partial \sigma} \quad \textit{and} \quad g_2(x,\sigma) = \frac{\partial g_1(x,\sigma)}{\partial \sigma}.$$

Example 2.1 (GED instrumental density) Consider the case in which h belongs to the class of the Generalized Error Distributions of shape parameter $\kappa > 0$, defined by

$$h_\kappa(x) = \frac{\kappa}{\Gamma(1/\kappa)2^{1+1/\kappa}} e^{-\frac{|x|^\kappa}{2}},$$

which will be denoted by GED(κ). We then have, for $x \neq 0$,

$$\frac{h'}{h}(x) = -\frac{\kappa |x|^\kappa}{2x}.$$

In view of (2.8), we obtain

$$\sigma_* = \left(\frac{\kappa}{2} E|\eta_1|^\kappa\right)^{1/\kappa}.$$

By (2.18) and (2.22) given in the proof of Lemma 2.1,

$$g_1\left(\frac{\eta_1}{\sigma_*}, 1\right) = -1 + \frac{|\eta_1|^\kappa}{E\left|\eta_1\right|^\kappa}, \quad g_2\left(\frac{\eta_1}{\sigma_*}, 1\right) = 1 - (1+\kappa)\frac{|\eta_1|^\kappa}{E\left|\eta_1\right|^\kappa}$$

and

$$\tau_h := \tau_{GED} = \frac{4}{\kappa^2}\left(\frac{E\left|\eta_1\right|^{2\kappa}}{\left(E\left|\eta_1\right|^\kappa\right)^2} - 1\right). \tag{2.10}$$

To give a more explicit example, assume that we have a standard GARCH(1,1) with parameter $\theta_0 = (\omega_0, \alpha_0, \beta_0)$ and $\eta_t \sim \mathcal{N}(0,1)$. For this distribution we have $E|\eta_1| = \sqrt{2/\pi}$. If we take the double exponential distribution $(1/4)e^{-|x|/2}$ as instrumental density h, which corresponds to the GED(1) , then $\hat{\theta}_n^*$ converges to $\theta_0^* = (2\omega_0/\pi, 2\alpha_0/\pi, \beta_0)$. Moreover the asymptotic variance is obtained with $\tau_h = 2\pi - 4$.

Example 2.2 (Double Generalized Gamma instrumental density) Now consider a larger class of densities, which contains, in particular, the GED, the Laplace, the double Weibull, Rayleigh and Maxwell, and the Gaussian distributions. Assume that h follows a double generalized Gamma (dgG) distribution $\Gamma(b, p, d)$ with parameters $b > 0$, $p > 0$ and $d > 0$, defined by the density

$$h(x) = h_{dgG}(x) = \frac{db^p}{2\Gamma(\frac{p}{d})}|x|^{p-1}e^{-|bx|^d}.$$

For $x \neq 0$, we have

$$x\frac{h'}{h}(x) = p - 1 - d|bx|^d.$$

In view of (2.8), we have $\sigma_* = \left(\frac{db^d}{p}E|\eta_1|^d\right)^{1/d}$. Thus,

$$g_1\left(\frac{\eta_1}{\sigma_*}, 1\right) = p\left(\frac{|\eta_1|^d}{E\,|\eta_1|^d} - 1\right), \quad g_2\left(\frac{\eta_1}{\sigma_*}, 1\right) = p\left(1 - (d+1)\frac{|\eta_1|^d}{E\,|\eta_1|^d}\right).$$

We then have

$$\tau_h = \tau_{dgG} = \frac{4}{d^2}\left(\frac{E\,|\eta_1|^{2d}}{\left(E\,|\eta_1|^d\right)^2} - 1\right).$$

Note that τ_{dgG} is equal to τ_{GED} when $\kappa = d$.

Therefore, compared to the GED, the introduction of the more complicated class of the dgG distributions is useless, because it does not lead to any efficiency gain.

Example 2.3 (Student instrumental density) Now consider the case where the instrumental density h is the Student distribution with ν degrees of freedom

$$h(x) = h_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

By (2.18) and (2.22), we have

$$g_1(x, \sigma) = \frac{\nu(x-\sigma)(x+\sigma)}{\sigma(x^2 + \nu\sigma^2)}, \quad g_2(x, \sigma) = -\frac{\nu\left\{x^4 + x^2(1+3\nu)\sigma^2 - \nu\sigma^4\right\}}{\sigma^2(x^2 + \nu\sigma^2)^2}.$$

In view of (2.8), the parameter σ_* satisfies

$$E\frac{\eta_1^2}{\nu\sigma_*^2+\eta_1^2}=\frac{1}{\nu+1}.$$

Contrary to what happens in Example 2.1, the parameters σ_* and τ_h do not have simple expressions as a function of ν and of the distribution of η_1, but can be obtained by numerical algorithms.

2.1.2 Estimating the VaR parameter

For the general volatility model (2.4), we have

$$\mathrm{VaR}_t(\alpha)=-\sigma_{t+1}(\theta_0^*)\xi_\alpha^*,$$

where ξ_α^* denotes the α-quantile of $\eta_t^*:=\eta_t/\sigma_*$. Note that, when $\xi_\alpha^*<0$, **A1** entails

$$\mathrm{VaR}_t(\alpha)=\sigma_{t+1}\left(\theta_{0,\alpha}\right)\quad\text{where}\quad\theta_{0,\alpha}=H\left(\theta_0^*,-\xi_\alpha^*\right).$$

The parameter $\theta_{0,\alpha}$ is called the VaR parameter in Francq and Zakoïan (2012). Note that $\xi_\alpha:=\sigma_*\xi_\alpha^*$ is the α-quantile of η_t. Thus we have $\theta_{0,\alpha}=H\left(\theta_0^*,-\xi_\alpha^*\right)=H\left(\theta_0,-\xi_\alpha\right)$.

Let $\hat{\xi}_{\alpha,n}^*$ be the empirical quantile of the residuals $\hat{\eta}_t^*:=\epsilon_t/\widetilde{\sigma}_t(\hat{\theta}_n^*)$ for $t=1,\dots,n$. We now give an intermediate result that will be used to obtain the asymptotic distribution of two-step estimators of the VaR parameter.

Theorem 2.1 *Assume η_1 has a density f, continuous at ξ_α, such as $f(\xi_\alpha)>0$. Under the assumptions of Lemma 2.1, and assuming that* **A5** *and* **A12** *hold with $\delta>1$, we have*

$$\sqrt{n}\begin{pmatrix}\hat{\theta}_n^*-\theta_0^*\\ \hat{\xi}_{\alpha,n}^*-\xi_\alpha^*\end{pmatrix}\stackrel{\mathcal{L}}{\to}\mathcal{N}\left\{0,\Sigma^*:=\begin{pmatrix}\Sigma_{11}^* & \Sigma_{12}^*\\ \Sigma_{12}^{*\,\prime} & \Sigma_{22}^*\end{pmatrix}\right\},$$

where

$$\begin{aligned}\Sigma_{11}^* &= \tau_h J_*^{-1},\\ \Sigma_{12}^* &= -\left\{\xi_\alpha^*\tau_h-\frac{4c_\alpha}{\sigma_*f(\xi_\alpha)Eg_2(\eta_0^*,1)}\right\}J_*^{-1}\Omega_*,\\ \Sigma_{22}^* &= \frac{\tau_h(\xi_\alpha^*)^2}{4}-\frac{2c_\alpha\xi_\alpha^*}{\sigma_*f(\xi_\alpha)Eg_2(\eta_0^*,1)}+\frac{\alpha(1-\alpha)}{\sigma_*^2f^2(\xi_\alpha)},\end{aligned}$$

with $\Omega_*=ED_t(\theta_0^*)$, $c_\alpha=Cov(\mathbf{1}_{\{\eta_t^*<\xi_\alpha^*\}},g_1(\eta_t^*,1))$.

In the case $h = \phi$ we retrieve Theorem 4.2 in Francq and Zakoïan (2012).

Note that $\hat{\theta}^*_{n,\alpha}$ converges to the VaR parameter $\theta_{0,\alpha}$. The star symbol is used to emphasize that, contrary to the parameter, the estimator depends on h.

The delta method immediately gives the following result.

Corollary 2.1 *Under the assumptions of Theorem 2.1 and if H is differentiable at $(\theta_0^*, -\xi_\alpha^*)$, with $\xi_\alpha^* < 0$, we have*

$$\sqrt{n}\left(\hat{\theta}^*_{n,\alpha} - \theta_{0,\alpha}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, G_* \Sigma^* G_*'\right),$$

where

$$G_* = \left[\frac{\partial H(\theta, K)}{\partial(\theta', K)}\right]_{(\theta_0^*, -\xi_\alpha^*)}.$$

By empirically estimating the asymptotic variance, this corollary makes it possible to obtain a confidence interval at an asymptotic statistical estimation-risk level α_1 for the risk parameter at the market-risk level α. Using again the delta method, confidence intervals for $\mathrm{VaR}_t(\alpha) = \sigma_{t+1}(\theta_{0,\alpha})$ at a given estimation-risk level can be deduced, exactly as Francq and Zakoïan (2012) did for the VaR estimation method based on the Gaussian QMLE.

The following result shows that the estimator of the VaR parameter is not sensitive to a scaling of the instrumental density.

Corollary 2.2 *Under the assumptions of Corollary 2.1, and if* **A1** *holds true when σ_t is replaced by $\tilde{\sigma}_t$,* i.e. *if*

$$K\tilde{\sigma}_t(\theta) = \tilde{\sigma}_t(\theta)\left\{H(\theta, K)\right\}, \tag{2.11}$$

then the estimator $\hat{\theta}^*_{n,\alpha}$ is not changed if $h(x)$ is replaced by $h_s(x) = s^{-1}h(s^{-1}x)$, for any $s > 0$.

In the standard GARCH(1,1) case, it is easy to see that (2.11) is satisfied when the initial value $\tilde{\sigma}_0(\theta)$ is chosen equal to zero.

2.2 Application to GARCH models

For particular GARCH models, we now verify the regularity conditions of Lemma 2.1, and we give a more explicit expression for the asymptotic variance of Corollary 2.1. We begin with the GARCH(1,1) model, and extend the result for a much wider class.

2.2.1 The first-order GARCH model

First begin with the GARCH(1,1) case, under Assumption **C**. In that case, the matrix G_* of Corollary 2.1 is given by

$$G_* = \begin{pmatrix} (\xi_\alpha^*)^2 & 0 & 0 & -2\xi_\alpha^*\omega_0^* \\ 0 & (\xi_\alpha^*)^2 & 0 & -2\xi_\alpha^*\alpha_0^* \\ 0 & 0 & 1 & 0 \end{pmatrix} := \left(A_* \quad -2\xi_\alpha^* \begin{pmatrix} \omega_0^* \\ \alpha_0^* \\ 0 \end{pmatrix} \right).$$

Note also that, for any $\theta_0^* = (\omega_0^*, \alpha_0^*, \beta_0^*) \in \Theta$, we have

$$\begin{aligned} (\omega_0^*, \alpha_0^*, 0)\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} &= \omega_0^* + \alpha_0^*\epsilon_{t-1}^2 + \beta_0^*\left\{(\omega_0^*, \alpha_0^*, 0)\frac{\partial\sigma_{t-1}^2(\theta_0^*)}{\partial\theta}\right\} \\ &= \sum_{i=0}^{\infty}\beta_0^{*i}\left\{\omega_0^* + \alpha_0^*\epsilon_{t-i}^2\right\} = \sigma_t^2(\theta_0^*). \end{aligned}$$

It follows that

$$\frac{1}{\sigma_t(\theta_0^*)}\frac{\partial\sigma_t(\theta_0^*)}{\partial\theta'}\begin{pmatrix} \omega_0^* \\ \alpha_0^* \\ 0 \end{pmatrix} = \frac{1}{2} \quad a.s.,$$

and thus

$$\Omega_*'\begin{pmatrix} \omega_0^* \\ \alpha_0^* \\ 0 \end{pmatrix} = \frac{1}{2}, \quad J_*\begin{pmatrix} \omega_0^* \\ \alpha_0^* \\ 0 \end{pmatrix} = 2\Omega_*, \quad J_*^{-1}\Omega_* = \frac{1}{2}\begin{pmatrix} \omega_0^* \\ \alpha_0^* \\ 0 \end{pmatrix}, \quad \Omega_*'J_*^{-1}\Omega_* = \frac{1}{4}.$$

The second equality of the previous line shows that

$$\mathrm{Var}\left(\frac{1}{\sigma_t^2(\theta_0^*)}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta}\right) = J_* - 4\Omega_*\Omega_*' = J_*(J_*^{-1} - \Psi_*)J_*,$$

where

$$\Psi_* = \begin{pmatrix} \omega_0^* \\ \alpha_0^* \\ 0 \end{pmatrix}\begin{pmatrix} \omega_0^* & \alpha_0^* & 0 \end{pmatrix} = \begin{pmatrix} \omega_0^{*2} & \omega_0^*\alpha_0^* & 0 \\ \omega_0^*\alpha_0^* & \alpha_0^{*2} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Under **A9**, which is implied by the identifiability condition in Assumption **C**, the matrix $\mathrm{Var}\left(\frac{1}{\sigma_t^2(\theta_0^*)}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta}\right)$ is positive definite. It follows that

$$J_*^{-1} - \Psi_* \quad \text{is positive definite.} \tag{2.12}$$

Moreover we have

$$\begin{aligned} G_*\Sigma^*G_*' &= \tau_h A_* J_*^{-1} A_* + \left(\frac{4(\xi_\alpha^*)^2\alpha(1-\alpha)}{\sigma_*^2 f^2(\xi_\alpha)} - \tau_h(\xi_\alpha^*)^4\right)\Psi_* \\ &= \tau_h A_*(J_*^{-1} - \Psi_*)A_* + \frac{4(\xi_\alpha^*)^2\alpha(1-\alpha)}{\sigma_*^2 f^2(\xi_\alpha)}\Psi_*. \end{aligned}$$

For the last equality we used that $A_*\Psi_*A_* = (\xi_\alpha^*)^4\Psi_*$.

Now we introduce analogs of the starred symbols, which are independent of the instrumental density h, using the matrix transformation

$$M_* = \begin{pmatrix} \frac{1}{\sigma_*^2}I_2 & 0_2 \\ 0_2' & 1 \end{pmatrix}.$$

We thus define $A = M_*^{-1}A_*$ and $\Psi = M_*\Psi_*M_* = \sigma_*{}^{-4}\Psi_*$. Note also that

$$\theta_0 \quad = \quad M_*\theta_0^*, \quad D_t(\theta_0^*) = M_*D_t(\theta_0) \quad \text{and} \quad J_* = M_*JM_*.$$

With this notation, we have

$$G_*\Sigma^*G_*{}' = \tau_h A(J^{-1} - \Psi)A + \frac{4\xi_\alpha^2\alpha(1-\alpha)}{f^2(\xi_\alpha)}\Psi. \tag{2.13}$$

The instrumental density h_1 is said to be more efficient than h_2, which is denoted by $h_1 \succ h_2$, if the difference of the asymptotic variances given by (2.13) is positive definite. In the asymptotic variance, only τ_h depends on h. In view of (2.12), this shows that $h_1 \succ h_2$ if and only if $\tau_{h_1} < \tau_{h_2}$.

2.2.2 The Asymmetric Power GARCH model

Ding, Granger and Engle (1993) introduced the so-called Asymmetric Power GARCH (APARCH) models, which include the standard GARCH of Bollerslev (1991), the TARCH of Zakoïan (1994), the GJR of Glosten, Jagannathan and Runkle (1993) and many other popular specifications of the volatility. Letting $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$, the model is defined by

$$\begin{cases} \epsilon_t = \sigma_t\eta_t \\ \sigma_t^\delta = \omega_0 + \sum_{i=1}^q \alpha_{0i+}(\epsilon_{t-i}^+)^\delta + \alpha_{0i-}(\epsilon_{t-i}^-)^\delta + \sum_{j=1}^p \beta_{0j}\sigma_{t-j}^\delta \end{cases} \tag{2.14}$$

where the coefficients satisfy $\alpha_{0i+} \geq 0$, $\alpha_{0i-} \geq 0$, $\beta_{0j} \geq 0$, $\omega_0 > 0$ and $\delta > 0$. The standard GARCH is obtained with $\delta = 2$ and $\alpha_{0i-} = \alpha_{0i+}$. When $\alpha_{0i-} > \alpha_{0i+}$, a negative return has a higher impact on the future volatility than a positive return of the same magnitude, which is a well-documented stylized fact that is called "leverage effect".

Hamadeh and Zakoïan (2011) showed that the power parameter δ is not easily estimated. We therefore consider that δ is fixed. In many applications, $\delta = 1$ (as in the TARCH) or $\delta = 2$ (as in the GJR model). As in Assumption **C**, let $\gamma(\theta_0)$ be the top-Lyapunov exponent associated with (2.14). Hamadeh and Zakoïan (2011) showed the CAN of the Gaussian QMLE of $\theta_0 = (\omega_0, \alpha_{01+}, \dots, \alpha_{0q-}, \beta_{01}, \dots, \beta_{0p})'$ under the assumption:

D: $\gamma(\theta_0) < 0$; θ_0 belongs to the interior of Θ; there exists $\underline{\omega} > 0$ such that, $\forall \theta \in \Theta$, $\omega > \underline{\omega}$ and $\sum_{j=1}^{p} \beta_j < 1$; the support of the distribution of η_1 contains at least 3 points; $P[\eta_t > 0] \in (0,1)$; if $p > 0$, $\mathcal{B}_{\theta_0}(z)$ has no common root with $\mathcal{A}_{\theta_0+}(z) = 1 - \sum_{i=1}^{q} \alpha_{0i+} z^i$ and $\mathcal{A}_{\theta_0-}(z) = 1 - \sum_{i=1}^{q} \alpha_{0i-} z^i$; $\mathcal{A}_{\theta_0+}(1) + \mathcal{A}_{\theta_0-}(1) \neq 0$ and $\alpha_{0q,+} + \alpha_{0q,-} + \beta_{0p} \neq 0$ (with the notation $\alpha_{00,+} = \alpha_{00,-} = \beta_{00} = 1$)

and under the identifiability condition $E\eta_1^2 = 1$ (that we do not assume in our framework).

The following theorem extends the results obtained in the previous section.

Theorem 2.2 *Consider the APARCH(p,q) model (2.14) under Assumption* **D**. *Assume η_1 has a density f, continuous at $\xi_\alpha < 0$, such as $f(\xi_\alpha) > 0$. If the instrumental density h satisfies* **A4**, **A5**, **A7**, **A8** *and* **A11**, *then the two-step estimator of the VaR parameter at the confidence level $\alpha \in (0,1)$ satisfies*

$$\sqrt{n}\left\{\hat{\theta}^*_{n,\alpha} - H\left(\theta_0^*, -\xi_\alpha^*\right)\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, G_*\Sigma^* G_*'\right),$$

where, for $\xi > 0$,

$$H\left(\omega, \alpha_{1+}, \dots, \alpha_{q-}, \beta_1, \dots, \beta_p, \xi\right) = \left(\xi^\delta \omega, \xi^\delta \alpha_{1+}, \dots, \xi^\delta \alpha_{q-}, \beta_1, \dots, \beta_p\right)$$

and

$$G_*\Sigma^* G_*' = \tau_h A(J^{-1} - \Psi)A + \frac{4\xi_\alpha^2 \alpha(1-\alpha)}{f^2(\xi_\alpha)}\Psi,$$

where $\overline{\theta}_0' = (\omega_0, \alpha_{01+}, \dots, \alpha_{0q-}, 0, \dots, 0)$,

$$A = \operatorname{diag}\left\{(-\xi_\alpha)^\delta I_{2q+1}, I_p\right\}, \quad \Psi = \overline{\theta}_0 \overline{\theta}_0', \quad J = 4ED_1(\theta_0)D_1'(\theta_0).$$

For the instrumental densities h_1 and h_2 , we have $h_1 \succ h_2$ if and only if $\tau_{h_1} < \tau_{h_2}$.

Remark 2.3 (On the optimal instrumental density) This theorem shows that an instrumental density h with the smallest value of τ_h is optimal. It is worth noting that the knowledge of the distribution of η_1, up to some (unknown) scaling constant, is sufficient to determine if h is optimal within the class of the two-step estimators introduced in this chapter. In particular the optimality of h: 1) does not depend on θ_0^*, or even on the volatility model; 2) does not depend on α.

Francq and Zakoïan (2013) compared the two-step estimator based on ϕ with a one step estimator. As in 1), the ranking of the two estimators is the same regardless of the model. However the relative efficiency of their two methods varies with α, which is not the case here.

Note also that the optimal instrumental density for estimating the VaR parameter is the same as that obtained by Fan et al. (2014) for their three-step estimator of the volatility parameter. This result is a little bit surprising because the present chapter and that of Fan et al. concern the estimation of two different parameters and do not use the same statistical techniques. In particular, contrary to Fan et al., we use quantile regression techniques for our estimator.

2.2.3 Optimal choice of the instrumental density

In view of Theorem 2.2, the optimal h (within a given class of instrumental densities satisfying the assumptions of the theorem) has the smallest τ_h. We first give an example of density h for which τ_h is a function of moments of η_1 that can be empirically estimated. We then give an example in which τ_h is not explicit, but can however be easily estimated.

GED instrumental distribution

Consider the case in which h is the GED(κ) distribution of Example 2.1. The value κ_0 of κ which minimizes (2.10) is considered as optimal for the parameter of a GED instrumental density. It is thus said that the GED(κ_0) is GED-optimal. Of course, for some distribution of η_t, it may exist another instrumental density, not belonging to the GED class, with a smaller τ_h than that of the GED(κ_0). In other words, the GED-optimal instrumental density is not always optimal in the strong sense. An empirical estimator of κ_0 can be obtained as follows. Let $\hat{\eta}_t = \epsilon_t/\tilde{\sigma}_t(\hat{\theta}_n)$, $t = 1, \dots, n$, be the residuals obtained from a first-step estimation procedure, which is consistent but not necessarily optimal, for example the Gaussian QMLE. An estimator of the parameter κ_0 for the GED-optimal instrumental density is defined by

$$\hat{\kappa} = \arg\min_{\kappa\in\mathcal{K}} \frac{1}{\kappa^2}\left(\frac{\hat{\mu}_{2\kappa}}{\hat{\mu}_\kappa^2} - 1\right), \quad \hat{\mu}_r = \frac{1}{n}\sum_{t=1}^n |\hat{\eta}_t|^r,$$

where $\mathcal{K}$ is a bounded interval containing κ_0. Note that it is important to minimize over a bounded interval because, by Lemma 3.1 in Francq et al. (2011), for any fixed n, we have

$$\frac{1}{\kappa^2}\left(\frac{\hat{\mu}_{2\kappa}}{\hat{\mu}_\kappa^2} - 1\right) \to 0, \qquad \text{as } \kappa \to \infty.$$

Student instrumental distribution

As in Example 2.3, let us take the Student distribution with ν degrees of freedom as instrumental density h. The parameters σ_* and τ_h can be estimated as follows. Let $\hat{\eta}_1, \dots, \hat{\eta}_n$ be the residuals of a first-step estimation procedure. Let C and S be compact subsets of $]0, \infty[$. For any value of $\nu \in C$, σ_* can be estimated by

$$\hat{\sigma}_* = \arg\max_{\sigma \in S} \sum_{t=1}^{n} g(\hat{\eta}_t, \sigma).$$

An estimator of the parameter of the "Student-optimal" instrumental density is then obtained as

$$\hat{\nu} = \arg\min_{\nu \in C} \frac{n^{-1} \sum_{t=1}^{n} g_1^2\left(\hat{\sigma}_*^{-1}\hat{\eta}_t, 1\right)}{\left\{n^{-1} \sum_{t=1}^{n} g_2\left(\hat{\sigma}_*^{-1}\hat{\eta}_t, 1\right)\right\}^2}. \tag{2.15}$$

2.2.4 Suboptimality of the naive adaptive approach

Assume a parametric form $h_\kappa(x)$, $\kappa \in \mathcal{K}$ for the instrumental density. We know that the optimal instrumental density is the (unknown) distribution f of η_1, or equivalently any scaled version $\sigma^{-1} f(x/\sigma)$, $\sigma > 0$, of this density (see Corollary 2.2). If some scaled version of f belongs to the chosen class of parametric instrumental densities, *i.e.* if $f(x) = \sigma_0^{-1} h_{\kappa_0}(x/\sigma_0)$ for some $\kappa_0 \in \mathcal{K}$ and some $\sigma_0 > 0$, then the optimal instrumental density can be found by the (quasi-)maximum likelihood procedure

$$(\hat{\kappa}, \hat{\sigma}) = \arg \max_{(\kappa,\sigma) \in \mathcal{K} \times (0,\infty)} \sum_{t=1}^{n} \log \sigma^{-1} h_\kappa(\hat{\eta}_t/\sigma),$$

where $\hat{\eta}_t = \epsilon_t / \tilde{\sigma}_t(\hat{\theta}_n)$, $t = 1, \dots, n$, are the residuals obtained from a Gaussian QMLE, or any other consistent first-step estimation procedure. Even if f does not belong to the class of densities, the procedure makes sense and converges, under general regularity conditions (see White 1982), to a minimizer of a Kullback-Leibler divergence, solution to

$$(\kappa^*, \sigma^*) = \arg \max_{(\kappa,\sigma) \in \mathcal{K} \times (0,\infty)} E \log \sigma^{-1} h_\kappa(\eta_1/\sigma).$$

For example, consider the case where the instrumental density belongs to the class of the GED(κ) distributions. We then have,

$$\sigma^* = \left(\frac{\kappa^* E|\eta_1|^{\kappa^*}}{2}\right)^{1/\kappa^*},$$

where

$$\kappa^* = \arg\max_{\kappa\in\mathcal{K}} \log\left(\frac{\kappa}{\Gamma(1/\kappa)2^{1+1/\kappa}}\right) - \frac{1}{\kappa}\left\{\log\left(\frac{\kappa E|\eta_1|^\kappa}{2}\right)+1\right\}.$$

Let GED(κ_0) be the GED-optimal instrumental density, and let τ_0 be the value of τ_h corresponding to this density, *i.e.* τ_0 is the smallest value of τ_h when h is GED(κ). In view of (2.10), we have

$$\tau_0 = \frac{4}{\kappa_0^2}\left(\frac{E\left|\eta_1\right|^{2\kappa_0}}{\left(E\left|\eta_1\right|^{\kappa_0}\right)^2}-1\right), \quad \kappa_0 = \arg\min_{\kappa}\frac{4}{\kappa^2}\left(\frac{E\left|\eta_1\right|^{2\kappa}}{\left(E\left|\eta_1\right|^{\kappa}\right)^2}-1\right).$$

Let τ^* be the value of τ_h when h is the GED(κ^*). This τ^* is optimal (*i.e.* minimal) when the density f of η_1 is a rescaled GED, and in this case we have $\tau^* = \tau_0$. In general, there is no guarantee that τ^* be optimal in the class of the GED instrumental density, *i.e.* that $\tau^* = \tau_0$.

2.2.5 Extension to other conditional risk measures

VaR is used by academics to define more sophisticated risk measures and VaR constitutes a powerful tool for professional risk managers, but it has been criticized for giving a too limited view of the actual risk level. In particular, VaR says nothing on what happens when losses exceed VaR. The expected shortfall (ES) is a popular alternative risk measure which circumvents this problem by measuring the average loss in the case of losses exceeding VaR. Another argument often given against VaR is that it does not satisfy the subadditivity property (see *e.g.* Artzner, Delbaen, Eber and Heath (1999), Wirch and Hardy (1999)). That means that the VaR of an average of risky assets can be larger than the average of the VaR of the individual assets.[2]

The ES satisfies the subadditivity property and constitues a leading example of the wide class of the Distortion Risk Measures (DRM) (see Wang (2000) and the references therein). A conditional DRM is defined by

$$\mathrm{DRM}_t = \int_0^1 \mathrm{VaR}_t(u)dG(u), \tag{2.16}$$

[2]That the risk of an average must be less than the average of the risks is however questionable. The usual central limit theorem (CLT) leads us to think that the answer should be positive, but this is not the case when considering generalized CLT's for variables without second order moments. Indeed, the risk of an average of iid Cauchy variables is the risk of a single Cauchy variable. More generally, an average of iid alpha-stable random variables with tail index smaller than 1 remains alpha-stable, but its scale increases, and thus the average should have a larger risk.

whenever the integral exists, where G is a cumulative distribution function (cdf) on $[0,1]$ that is called the distortion function. The DRM can be interpreted as a weighted sum of VaR's, where the weights are the increases of the distortion function. The ES is obtained with $G(u) = (u/\alpha)1_{[0,\alpha[}(u) + 1_{[\alpha,\infty[}(u)$. Other examples of DRM are the proportional hazard DRM, obtained with $G(u) = u^r$, and the exponential DRM, obtained with $G(u) = (1-e^{ru})/(1-e^r)$, $r > 0$. Assuming $\int_0^1 \xi_u dG(u) < 0$, under **A1** we have

$$\mathrm{DRM}_t = -\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) \int_0^1 \xi_u dG(u) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_{0,G}),$$

where

$$\theta_{0,G} = H\left(\theta_0, -\int_0^1 \xi_u dG(u)\right) \tag{2.17}$$

can be called the conditional DRM risk parameter. The assumption $\int_0^1 \xi_u dG(u) < 0$, which is required because the constant K in **A1** must be positive, is satisfied when the distortion function puts enough weight on extreme risks, corresponding to small u's and negative quantiles ξ_u. A natural estimator of the parameter $\theta_{0,G}$ is then

$$\hat{\theta}^*_{n,G} = H\left(\hat{\theta}^*_n, -\int_0^1 \hat{\xi}^*_{n,u} dG(u)\right).$$

For estimating the conditional VaR, $-\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0)\xi_u$, the optimal instrumental density h does not depend on u (see Remark 2.3). For estimating the weighted VaR, $\mathrm{DRM}_t = -\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) \int_0^1 \xi_u dG(u)$, it is natural to chose the same optimal instrumental density h, which minimizes τ_h, at least in the APARCH case (see Theorem 2.2).

2.3 Numerical illustrations

We first consider a theoretical framework in which the distribution of η_t is assumed to be known. Considering two classes of instrumental densities, the GED(κ) and the Student St_ν distributions, we determined the best instrumental densities within each class, and we compared them with the standard Gaussian density in terms of asymptotic relative efficiency. In the second subsection, Monte Carlo experiments are used to compare the finite sample performance of the different VaR estimation procedures. The last subsection proposes illustrations on financial series.

2.3.1 Theoretical comparison of the asymptotic efficiencies

Assume that η_1 follows the double generalized Gamma distribution $\Gamma(b,p,d)$ considered in Example 2.2. We then have $E|\eta_1|^r = b^{-r}\Gamma((p+r)/d)/\Gamma(p/d)$. In view of (2.10), the minimal value of τ_h, which is obtained when $h \sim \Gamma(b, p, d)$, is given by

$$\tau_{opt} = \frac{4}{pd}.$$

With the standard approach based on the Gaussian QMLE, we have

$$\tau_\phi = \left(\frac{E\left|\eta_1\right|^4}{\left(E\left|\eta_1\right|^2\right)^2} - 1 \right) = \left(\frac{\Gamma\left(\frac{p}{d}\right)\Gamma\left(\frac{p+4}{d}\right)}{\left\{\Gamma\left(\frac{p+2}{d}\right)\right\}^2} - 1 \right).$$

The asymptotic relative efficiency (ARE) of the generalized QMLE based on the instrumental density h with respect to the standard Gaussian QMLE can be measured by the ratio

$$\text{ARE} = \frac{\tau_\phi}{\tau_h}.$$

In view of (2.10), the method based on the instrumental density GED(κ) is optimal (*i.e.* $\tau_{\text{GED}(\kappa)} = \tau_{opt}$) when $\kappa = d$. Figure 2.1 shows that, even if the instrumental densities GED(d) and $\Gamma(b, p, d)$ are asymptotically equivalent, they can be surprisingly different.

Figure 2.2 shows that the GED instrumental density can be much more efficient than the Gaussian one (indeed its ARE is much greater than 1 when d is small). The ARE reaches 1 for $d = \kappa = 2$. This was expected because the GED(2) and $\Gamma(\sqrt{1/2}, 1, 2)$ distributions both coincide with the standard Gaussian distribution. This figure also displays the ARE of the best Student instrumental density with respect to the Gaussian distribution. Even if the Student is generally not optimal (*i.e.* the Student-optimal is not optimal in the strong sense) when $\eta_t \sim \Gamma(b, p, d)$, it can also be much more efficient than the gaussian.

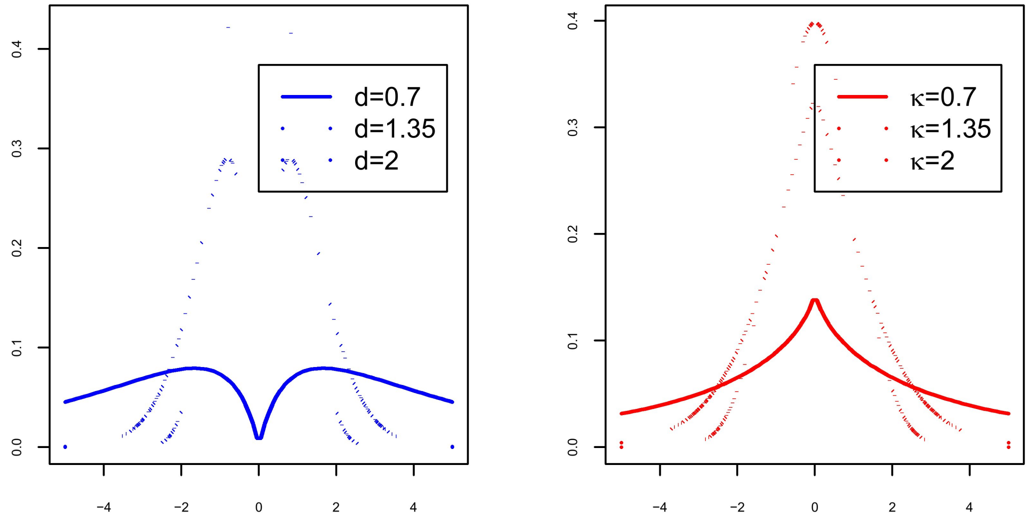


Figure 2.1 – *Density* $\Gamma(1, 2, d)$ *for* $d = 0.7$, $d = 1.35$ *and* $d = 2$ *(left panel) and density GED(*κ*) for* $\kappa = 0.7$, $\kappa = 1.35$ *and* $\kappa = 2$ *(right panel). The asymptotic distribution of the generalized QMLE based on* $\Gamma(b, p, d)$ *is the same as that based on GED(*κ*) when* $\kappa = d$.

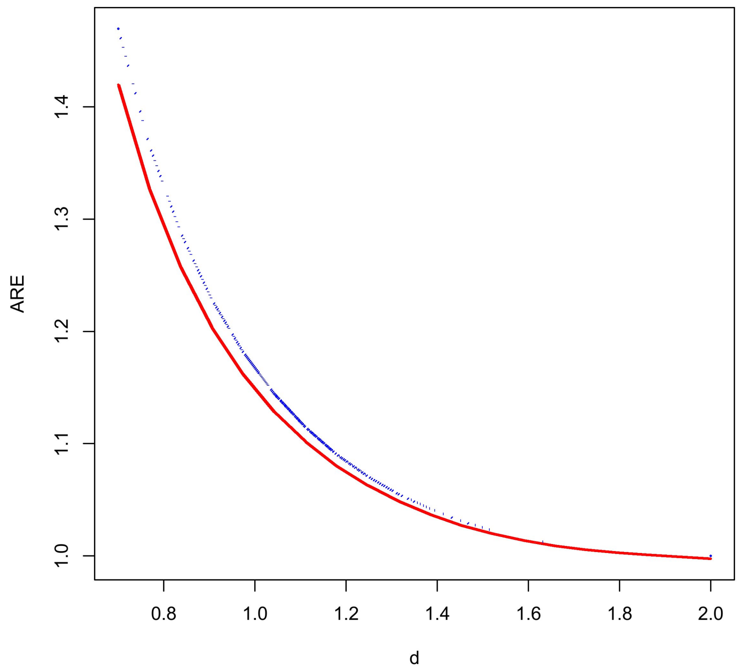


Figure 2.2 – *ARE of the generalized QMLE based on the optimal GED (dotted line), or based on the optimal Student instrumental density (full line), with respect to the Gaussian QMLE, when* $\eta_t \sim \Gamma(1, 2, d)$ *and* d *varies from* $d = 0.7$ *to* $d = 2$.

2.3.2 Simulation experiments

In the previous section, the selection of the GED-optimal or Student-optimal instrumental density is accomplished by assuming that the distribution of η_t is known, which is obviously unrealistic in practice. In this section, we first study if the selection of the optimal procedure can be satisfactorily done by using the estimated residuals. We thus simulate $N = 100$ independent trajectories of size $n = 1,000$ of a GARCH(1,1) model with $\theta_0 = (0.02, 0.002, 0.8)$ and $\eta_t \sim \Gamma(1, 2, d)$, where d takes 20 values between $d = 0.7$ and $d = 2$, as in Figure 2.2. For each simulation and each value of d, the parameter τ_ϕ is estimated by

$$\hat{\tau}_\phi = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 1, \qquad \hat{\mu}_r = \frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_t^r, \quad \hat{\eta}_t = \frac{\epsilon_t}{\widetilde{\sigma}_t(\widehat{\theta})},$$

where $\widehat{\theta}$ denotes the Gaussian QMLE. We then obtain an estimate of the optimal value of τ_{GED} by taking the minimum of

$$\frac{4}{\kappa^2}\left(\frac{\hat{\mu}_{2\kappa}}{\hat{\mu}_\kappa^2} - 1\right)$$

over $\kappa \in [0.1, 5]$. An estimate of the optimal value of τ_{St} is similarly obtained from (2.15). The curves of Figure 2.3 correspond to the average estimated ARE's over the N replications. The curves have very similar shapes to those of Figure 2.2, and lead to the same ranking of the estimation methods. This shows that one can actually select the asymptotically optimal method by choosing the method which minimizes the estimated value of τ computed from the residuals. Table 2.1 compares the actual accuracies of the different methods for estimating the VaR parameter at the 5% risk level. For clarity reasons, the results are only given for the 4 values of $d \in \{0.7, 0.97, 1.66, 2\}$. The column "bias" gives the average of the N estimation errors of the VaR parameter. The column RMSE gives the root mean square error of estimation. As expected from the asymptotic results (see Figure 2.2), the estimators based on the GED and Student instrumental densities are always very close, and they are much more efficient than the usual two-step estimator based on the Gaussian QMLE when the density of η_t is far from the Gaussian (*i.e.* when $d = 0.7$ or $d = 0.97$), whereas all the estimators are equivalent when d is close to 2 (which corresponds to the Gaussian case). Table 2.2 shows that, as expected from the theory, the ranking of the method is the same for the risk level of 1%. Note that there exist large biases, which can be explained by the boundary constraints of the parameter space. More precisely, for few simulations the conditional heteroscedasticity turns out not to be clearly visible on the generated series. In such a case, the estimated value of β is close to zero and, as a compensation, that of ω is large, which entails a large negative bias for the estimator of β and a huge positive bias for that of ω. Since $\hat{\beta}$ is constrained to be less than 1 and $\hat{\omega}$ is constrained to be positive, very large positive values of $\hat{\beta} - \beta_0$ and negative values of $\hat{\omega} - \omega_0$ are not possible. Note also that the biases and RMSE's vary a lot with d, which is not surprising because the variance of η_t is respectively equal to 42.2, 6.8, 1.4 and 1 when $d = 0.7, 0.97, 1.66$ and 2, and thus the scales of the simulated trajectories vary a lot. Of course, we have checked that the biases and RMSE's diminish when the sample size increases.

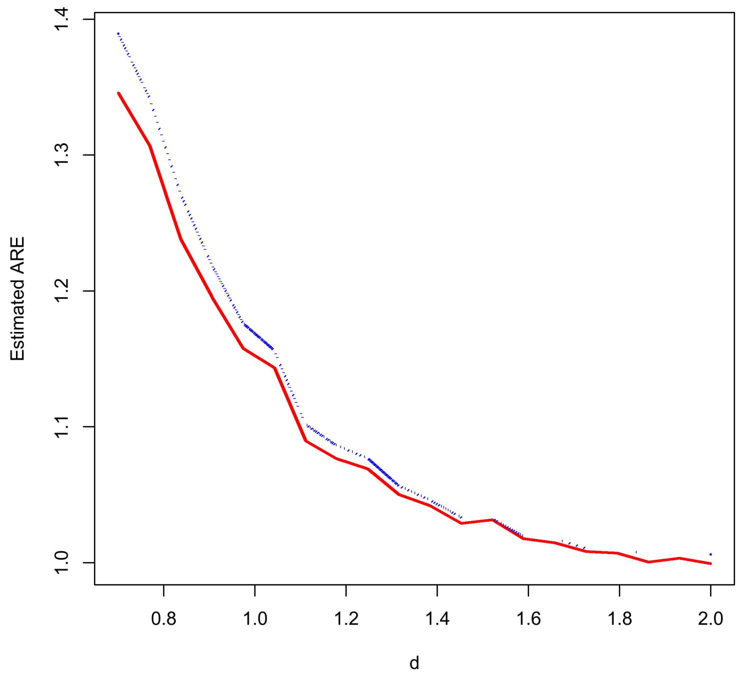


Figure 2.3 – *As figure 2.2, but the ARE's are estimated from the residuals of a GARCH(1,1) with innovations* $\eta_t \sim \Gamma(1, 2, d)$.

The Monte Carlo experiments displayed in this section are in perfect agreement with the asymptotic theory used in Section 2.3.1. As expected from Figure 2.2 an estimator of the VaR parameter based on a GED or Student generalized QMLE can be much more accurate than the usual estimator based on the Gaussian QMLE.

Table 2.1 – *Distribution of the estimation errors for the 5%-VaR parameter of a GARCH(1,1) model with $\eta_t \sim \Gamma(1, 2, d)$, using the standard Gaussian QMLE, the generalized QMLE based on the optimal GED instrumental density, or that based on the Student density.*

	Gaussian-QMLE		GED-QMLE		Student-QMLE	
VaR parameter ω						
d	bias	RMSE	bias	RMSE	bias	RMSE
0.7	2.211	4.845	0.735	2.658	1.170	3.531
0.97	1.129	1.261	0.445	0.807	0.667	0.901
1.66	0.053	0.118	0.052	0.116	0.054	0.117
2	0.047	0.088	0.044	0.085	0.049	0.089
VaR parameter α						
d	bias	RMSE	bias	RMSE	bias	RMSE
0.7	0.004	0.104	0.011	0.086	0.006	0.091
0.97	0.035	0.084	0.034	0.077	0.036	0.081
1.66	0.025	0.052	0.024	0.051	0.024	0.051
2	0.019	0.043	0.020	0.044	0.020	0.044
VaR parameter β						
d	bias	RMSE	bias	RMSE	bias	RMSE
0.7	-0.128	0.276	-0.041	0.149	-0.064	0.194
0.97	-0.625	0.686	-0.245	0.429	-0.370	0.486
1.66	-0.169	0.372	-0.165	0.361	-0.172	0.364
2	-0.214	0.388	-0.201	0.374	-0.224	0.396

2.3.3 Application to daily stock indices

We now consider the estimation of the VaR parameter for daily returns of 7 world stock market indices: CAC, DAX, FTSE, Nikkei, SMI (Swiss Market Index), SP500 and TSX (Toronto Stock Exchange). The data set comes from Yahoo Finance and covers the period from early January 1990 to the end of June 2013, when these historical data exist. The number of observations varies from 5721 (for the DAX) to 5934 (for FTSE).

For each series of log-returns ϵ_t, we estimated the VaR parameter $\theta_{0,\alpha}$ of GARCH(1, 1) models. Tables 2.3 and 2.4 report the estimated VaR parameters, their related standard deviations and the estimated τ_h's for three different instrumental densities h, namely the Gaussian, Student(ν) and GED(κ) distributions. For the last two instrumental densities, we chose the parameters ν and κ which

Table 2.2 – *As Table 2.1, but for the 1% risk level.*

	Gaussian-QMLE		GED-QMLE		Student-QMLE	
VaR parameter ω						
d	bias	RMSE	bias	RMSE	bias	RMSE
0.7	5.866	13.281	1.701	6.285	2.997	9.550
0.97	2.622	2.949	0.996	1.796	1.523	2.052
1.66	0.094	0.217	0.091	0.208	0.096	0.212
2	0.079	0.148	0.074	0.144	0.082	0.151
VaR parameter α						
d	bias	RMSE	bias	RMSE	bias	RMSE
0.7	-0.004	0.292	0.012	0.235	0.000	0.248
0.97	0.081	0.197	0.076	0.176	0.082	0.188
1.66	0.044	0.095	0.043	0.092	0.043	0.092
2	0.031	0.072	0.033	0.074	0.033	0.074
VaR parameter β						
d	bias	RMSE	bias	RMSE	bias	RMSE
0.7	-0.128	0.276	-0.041	0.149	-0.064	0.194
0.97	-0.625	0.686	-0.245	0.429	-0.370	0.486
1.66	-0.169	0.372	-0.165	0.361	-0.172	0.364
2	-0.214	0.388	-0.201	0.374	-0.224	0.396

minimize the τ_h's that are estimated from the QMLE residuals (as explained in Section 2.3.2). The estimated values of the τ_h's are thus the same for $\alpha = 5\%$ and $\alpha = 1\%$, which is in concordance with the asymptotic theory, since the τ_h's do not depend on α, nor on the volatility parameter θ_0. Recall that the most accurate estimator is that with the smallest τ_h. Therefore, the estimators based on the GED and Student distributions should be much more accurate than that based on the Gaussian density. This is not surprising because the Student and GED laws can have thicker tails than the normal distribution, and the financial series are known to have Leptokurtic conditional distributions. Thus, we addressed the issue of Leptokurticity through the use of Student and GED distributions. Over the 7 indices, it is clear to note that $\hat{\theta}^*_{n,\alpha}$ based on the GED and Student distributions are quite similar, with always a slight advantage (*i.e.* a smaller estimated τ_h) for the Student. The same conclusion can be drawn by looking at the estimated standard deviations, which are almost equal for the GED and Student distributions, and are clearly larger for the Gaussian instrumental density.

The adequacy of a VaR evaluation method is generally assessed by means of

Table 2.3 – *Comparison of estimators of the 5% level VaR parameter for 7 daily stock market returns. The estimated standard deviation are displayed in brackets.*

Index	h	$\omega_{5\%}$	$\alpha_{5\%}$	$\beta_{5\%}$	τ_h
CAC	ϕ	0.091 (0.021)	0.247 (0.030)	0.899 (0.011)	3.711
	GED	0.071 (0.015)	0.221 (0.024)	0.912 (0.008)	2.699
	St	0.065 (0.014)	0.220 (0.023)	0.914 (0.008)	2.537
DAX	ϕ	0.089 (0.026)	0.231 (0.041)	0.902 (0.016)	7.707
	GED	0.048 (0.011)	0.225 (0.024)	0.914 (0.008)	2.952
	St	0.045 (0.011)	0.230 (0.023)	0.913 (0.008)	2.676
FTSE	ϕ	0.037 (0.008)	0.243 (0.025)	0.906 (0.009)	2.780
	GED	0.035 (0.007)	0.230 (0.023)	0.911 (0.008)	2.513
	St	0.033 (0.007)	0.231 (0.023)	0.911 (0.008)	2.454
Nikkei	ϕ	0.153 (0.031)	0.286 (0.034)	0.878 (0.013)	3.517
	GED	0.110 (0.022)	0.249 (0.026)	0.897 (0.010)	2.803
	St	0.103 (0.020)	0.246 (0.025)	0.900 (0.009)	2.659
SMI	ϕ	0.137 (0.033)	0.353 (0.058)	0.845 (0.023)	7.429
	GED	0.076 (0.014)	0.319 (0.033)	0.877 (0.011)	2.908
	St	0.073 (0.013)	0.321 (0.032)	0.878 (0.010)	2.659
SP500	ϕ	0.028 (0.007)	0.204 (0.024)	0.918 (0.009)	3.777
	GED	0.020 (0.005)	0.192 (0.020)	0.926 (0.007)	2.997
	St	0.019 (0.005)	0.188 (0.019)	0.928 (0.007)	2.890
TSX	ϕ	0.021 (0.006)	0.230 (0.028)	0.914 (0.010)	4.347
	GED	0.016 (0.004)	0.204 (0.021)	0.924 (0.007)	2.887
	St	0.017 (0.004)	0.207 (0.021)	0.923 (0.007)	2.735

Table 2.4 – *As Table 2.3, but for the risk level 1%.*

Index	h	$\omega_{1\%}$	$\alpha_{1\%}$	$\beta_{1\%}$	τ_h
CAC	ϕ	0.198 (0.045)	0.537 (0.067)	0.899 (0.011)	3.711
	GED	0.153 (0.032)	0.478 (0.053)	0.912 (0.008)	2.699
	St	0.140 (0.030)	0.474 (0.050)	0.914 (0.008)	2.537
DAX	ϕ	0.203 (0.059)	0.525 (0.093)	0.902 (0.016)	7.707
	GED	0.112 (0.026)	0.526 (0.057)	0.914 (0.008)	2.952
	St	0.107 (0.024)	0.540 (0.056)	0.913 (0.008)	2.676
FTSE	ϕ	0.084 (0.018)	0.550 (0.061)	0.906 (0.009)	2.780
	GED	0.079 (0.017)	0.523 (0.057)	0.911 (0.008)	2.513
	St	0.076 (0.016)	0.528 (0.057)	0.911 (0.008)	2.454
Nikkei	ϕ	0.332 (0.068)	0.622 (0.076)	0.878 (0.013)	3.517
	GED	0.234 (0.047)	0.528 (0.058)	0.897 (0.010)	2.803
	St	0.221 (0.044)	0.527 (0.056)	0.900 (0.009)	2.659
SMI	ϕ	0.316 (0.077)	0.814 (0.137)	0.845 (0.023)	7.429
	GED	0.174 (0.032)	0.726 (0.079)	0.877 (0.011)	2.908
	St	0.165 (0.030)	0.729 (0.075)	0.878 (0.010)	2.659
SP500	ϕ	0.070 (0.017)	0.506 (0.061)	0.918 (0.009)	3.777
	GED	0.048 (0.012)	0.462 (0.05)	0.926 (0.007)	2.997
	St	0.047 (0.012)	0.455 (0.049)	0.928 (0.007)	2.890
TSX	ϕ	0.056 (0.015)	0.602 (0.083)	0.914 (0.01)	4.347
	GED	0.043 (0.011)	0.541 (0.064)	0.924 (0.007)	2.887
	St	0.044 (0.011)	0.546 (0.064)	0.923 (0.007)	2.735

out-of-sample prediction exercises, known under the name of backtesting procedures. To evaluate and compare the performance of the different VaR estimators, we implemented three standard backtests: the unconditional coverage (UC) test that the probability of violation (*i.e.* the observed loss exceeds the predicted VaR) is equal to the nominal level α, the independence (IND) test that the violations are independent, and the conditional coverage (CC) test (see Christoffersen (2003) for details). We illustrated the backtest procedures on the CAC index over the period 24/10/2008 to 01/07/2013 ($n = 1200$ observations), for three VaR estimators: the standard method based on the Gaussian QMLE, and the alternative methods based on the Student and GED optimal instrumental densities. For the period P1, we used the first 350 observations as a training set for estimating the VaR parameter by the 3 methods, and we used the next 150 observations for backtesting. We have repeated the exercise for two other periods: for P2 (resp. P3), the estimation period spans from observations 351 to 700 (resp. 701 to 1050) and the validation period spans from 701 to 850 (resp. 1051 to 1200). Table 2.5 displays the number of violations and the p-values of the tests for the 3 subperiods. The VaR evaluation procedure based on the Gaussian QMLE is clearly rejected in the first period for both levels 5% and 1%, whereas the methods based on the optimal GED and optimal Student distributions are never rejected. As emphasized by Escanciano and Olmo (2010), a VaR evaluation procedure can be rejected for two reasons: 1) the assumed parametric model is not correct (here the volatility model is not a GARCH(1,1)), 2) the parameter is not sufficiently well estimated. These two reasons correspond respectively to the model risk and to the estimation risk. The backtesting procedures are designed to detect the model risk, but do not incorporate the estimation risk. From the output of Table 2.5, the GARCH(1,1) model is plausible (otherwise the VaR estimations of the two other methods, which are also based on a GARCH(1,1), would also be rejected) but the estimation risk induced by the standard method is certainly higher than for the two other methods, which would explain that the backtests reject only for the first method.

2.4 Conclusion

In this chapter, we considered a general volatility model with an unknown volatility parameter θ_0, and an unknown distribution P_η for the iid noise. We did not make any identifiability assumption, such as $E\eta_t^2 = 1$, and we considered a generalized QMLE based on an arbitrary instrumental density h, which is not necessarily the density of P_η. We have shown that, under mild regularity conditions, the gQMLE converges however to some "pseudo-true" value θ_0^* which depends on θ_0 and on

Table 2.5 – *p-values of the three backtests for the levels 5% and 1%*

CAC	5%	Viol	UC	IND	CC	1%	Viol	UC	IND	CC
ϕ	P1	17	**0.00**	0.95	**0.01**	P1	7	**0.00**	0.41	**0.00**
	P2	9	0.59	0.55	0.72	P2	2	0.70	0.82	0.90
	P3	3	0.06	0.72	0.15	P3	1	0.66	0.91	0.90
GED	P1	9	0.59	0.28	0.48	P1	3	0.28	0.73	0.52
	P2	8	0.85	0.42	0.71	P2	3	0.28	0.73	0.52
	P3	3	0.06	0.73	0.15	P3	1	0.66	0.91	0.90
STD	P1	3	0.06	0.73	0.15	P1	1	0.66	0.91	0.90
	P2	8	0.85	0.42	0.71	P2	3	0.28	0.73	0.52
	P3	3	0.06	0.73	0.15	P3	1	0.66	0.91	0.90

some scale parameter depending on P_η and h.

Simply noting that, for any reasonable ARCH-type model, the ratio $\sigma_t(\theta_0^*)/\sigma_t(\theta_0)$ is constant, the conditional VaR at the level α can be obtained by multiplying $\sigma_t(\theta_0^*)$ by the α-quantile of $\eta_t^* = \epsilon_t/\sigma_t(\theta^*)$. This shows that the natural two-step method leads to a consistent estimation of the VaR, even if the instrumental density h does not coincide with P_η. The asymptotic and finite-sample accuracy of the method however depends on θ_0, h and P_η. We have shown that, for a large class of standard GARCH models, the optimal choice of h only depends on P_η and can be estimated easily. It is shown that, compared to the usual two-step method based on the Gaussian QMLE, important efficiency gains can be achieved by appropriately choosing the instrumental density.

Future extensions of this work could be the following. Firstly, it could be interesting to extend Corollary 2.1 in the case of a DRM parameter of the form (2.17). Such a result could be used to obtain confidence intervals for DRM that would integrate the estimation risk. This extension is however far from being trivial because it should involve the limit distribution of the random function $\sqrt{n}\left(\hat{\theta}_{n,\alpha}^* - \theta_{0,\alpha}\right)$ where α varies in $[0,1]$. Another potential extension would be to consider conditional risk measures for a time horizon larger than 1. Existing techniques are based on scenario simulations. The question of interest would be to determine whether such simulation techniques are more efficient at any horizon when they are based on models estimated by an optimal gQMLE than when they are based on the Gaussian QMLE. Finally, following Escanciano and Olmo (2010), it would be worth developing backtesting procedures which incorporate the estimation risk of the VaR parameter for the two-step procedures considered here.

2.5 Proofs

2.5.1 Proof of Lemma 2.1

The proof is similar to that of Theorem 2.1 in FZ. It rests on the following intermediate results:

$$i) \ \lim_{n\to\infty} \sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| = 0 , \quad a.s.$$

$$ii) \text{ if } \theta \neq \theta_0^* , \quad \mathbb{E}g(\epsilon_1, \sigma_1(\theta)) < \mathbb{E}g(\epsilon_1, \sigma_1(\theta_0^*)) ,$$

$iii)$ any $\theta \neq \theta_0^*$ has a neighborhood $V(\theta)$ such that

$$\limsup_{n\to\infty} \sup_{\theta^*\in V(\theta)} \widetilde{Q}_n(\theta^*) < \limsup_{n\to\infty} \widetilde{Q}_n(\theta_0^*) , \quad a.s.$$

where

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^n g(\epsilon_t, \sigma_t(\theta)),$$

$$iv) \ \lim_{n\to\infty} \sqrt{n} \sup_{\theta\in V(\theta_0^*)} \left\| \frac{\partial}{\partial\theta} Q_n(\theta) - \frac{\partial}{\partial\theta}\widetilde{Q}_n(\theta) \right\| = 0 , \quad \text{in probability,}$$

for some neighborhood $V(\theta_0^*)$ of θ_0^*,

$$v) \ J_* \text{ invertible and } \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta^*) \to \frac{Eg_2(\sigma_*^{-1}\eta_0, 1)}{4} J_* , \quad \text{in probability,}$$

for any θ^* between $\hat{\theta}_n^*$ and θ_0^*,

$$vi) \ \sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0^*) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{Eg_1^2(\sigma_*^{-1}\eta_0, 1)}{4} J_*\right).$$

We begin to show *i)*. First note that a Taylor expansion and **A5** show that

$$g(\epsilon_t, \widetilde{\sigma}_t(\theta)) - g(\epsilon_t, \sigma_t(\theta)) = g_1(\epsilon_t, \sigma_t^*(\theta)) \left\{\widetilde{\sigma}_t(\theta) - \sigma_t(\theta)\right\}$$

where

$$g_1(\epsilon, \sigma) = -\frac{1}{\sigma}\left\{1 + \frac{\epsilon}{\sigma}\frac{h'}{h}\left(\frac{\epsilon}{\sigma}\right) 1_{\epsilon\neq 0}\right\} \tag{2.18}$$

and $\sigma_t^*(\theta)$ is between $\widetilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. Using **A3** and **A5**, we then have almost surely

$$\begin{aligned}
\sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| &\leq C_1 n^{-1} \sum_{t=1}^n \sup_{\theta\in\Theta} |g_1(\epsilon_t, \sigma_t^*(\theta))| \rho^t \\
&\leq \frac{C_1}{n\underline{\omega}} \sum_{t=1}^n \rho^t \left\{1 + C_0\left(1 + \left|\frac{\epsilon_t}{\underline{\omega}}\right|^\delta\right)\right\}.
\end{aligned}$$

The Markov inequality and **A2** entail

$$\sum_{t=1}^{\infty} \mathbb{P}(\rho^t |\epsilon_t|^\delta > \varepsilon) \leq \sum_{t=1}^{\infty} \frac{\rho^{st/\delta} \mathbb{E}|\epsilon_t|^s}{\varepsilon^s} < \infty \tag{2.19}$$

and thus the proof of *i)* is completed by the Borel-Cantelli lemma.

To prove *ii)*, first note that by **A1** and (2.7) we have

$$\frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta_0)} = \sigma_*,$$

where σ_* is defined in **A4**. In view of **A3**-**A4** and (2.6), we thus have

$$\mathbb{E}\{g(\epsilon_1, \sigma_1(\theta)) - g(\epsilon_1, \sigma_1(\theta_0^*))\} = \mathbb{E}\left\{ g\left(\eta_t, \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}\right) - g(\eta_t, \sigma_*) \right\} \leq 0,$$

with equality if and only if $\theta = \theta_0^*$, which shows *ii)*.

We now turn to the proof of *iii)*. For any $\theta \in \Theta$ and any positive integer k, let $V_k(\theta)$ be the open ball with center θ and radius $1/k$. We have,

$$\begin{aligned}
& \limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} \widetilde{Q}_n(\theta^*) \\
\leq & \limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} Q_n(\theta^*) + \limsup_{n\to\infty} \sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| \\
\leq & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) \quad a.s.
\end{aligned}$$

where the second inequality comes from *i)*. Note that since h is integrable and continuous, h is bounded by some constant C. It follows, by **A3**, that

$$\mathbb{E} \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) < \log \frac{1}{\underline{\omega}} + \log C < \infty. \tag{2.20}$$

Using **A2** and an ergodic theorem for stationary and ergodic processes (X_t) such that $\mathbb{E}(X_t)$ exists in $\mathbb{R} \cup \{-\infty\}$ (see Billingsley, 1995, p. 284 and 495), it follows that

$$\limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} \widetilde{Q}_n(\theta^*) \leq \mathbb{E}X_{t,k}(\theta), \qquad X_{t,k}(\theta) = \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) \ .$$

When k tends to infinity, the sequence $\{X_{t,k}(\theta)\}_k$ decreases to $X_t(\theta) = g(\epsilon_t, \sigma_t(\theta))$. Thus $\{X_{t,k}^-(\theta)\}_k$ increases to $X_t^-(\theta)$. By the Beppo-Levi theorem, $\mathbb{E}X_{t,k}^-(\theta) \uparrow \mathbb{E}_{\theta_0} X_t^-(\theta)$ when $k \uparrow +\infty$. By (2.20), the fact that the sequence

$\{X_{t,k}^+(\theta)\}_k$ is decreasing, and the Lebesgue theorem, $\mathbb{E}X_{t,k}^+(\theta) \downarrow \mathbb{E}X_t^+(\theta)$ when $k \uparrow +\infty$. Thus we have shown that $\mathbb{E}X_{t,k}$ converges to $\mathbb{E}\{X_t(\theta)\}$ when $k \to \infty$. By *ii)*, *iii)* is proved.

The consistency is a consequence of **A7**, a standard compactness argument and of the intermediate results *i)-iii)*.

Now we prove *iv)*. We have

$$\begin{aligned}
\frac{\partial}{\partial\theta} Q_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n} g_1(\epsilon_t, \sigma_t(\theta)) \frac{\partial \sigma_t(\theta)}{\partial\theta}, \\
\frac{\partial}{\partial\theta} \widetilde{Q}_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n} g_1(\epsilon_t, \widetilde{\sigma}_t(\theta)) \frac{\partial \widetilde{\sigma}_t(\theta)}{\partial\theta}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sup_{\theta\in V(\theta_0^*)} \sqrt{n} \left\| \frac{\partial}{\partial\theta} Q_n(\theta) - \frac{\partial}{\partial\theta} \widetilde{Q}_n(\theta) \right\| \\
\leq\ & \sup_{\theta\in V(\theta_0^*)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |g_1(\epsilon_t, \sigma_t(\theta)) - g_1(\epsilon_t, \widetilde{\sigma}_t(\theta))| \left\| \frac{\partial \sigma_t(\theta)}{\partial\theta} \right\| \\
&+ \sup_{\theta\in V(\theta_0^*)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |g_1(\epsilon_t, \widetilde{\sigma}_t(\theta))| \left\| \frac{\partial \sigma_t(\theta)}{\partial\theta} - \frac{\partial \widetilde{\sigma}_t(\theta)}{\partial\theta} \right\|. \qquad (2.21)
\end{aligned}$$

In view of **A5** and **A10**, the last term is bounded by

$$\frac{C_1}{\sqrt{n}\underline{\omega}} \sum_{t=1}^{n} \rho^t \left\{ 1 + C_0 \left(1 + \left| \frac{\epsilon_t}{\underline{\omega}} \right|^{\delta} \right) \right\},$$

which is a.s. an $O(1/\sqrt{n})$ by arguments used to show *i)*. Thus it remains to show that the first term on the right-hand side of the inequality (2.21) converges also to zero a.s. as n tends to infinity. Noting that

$$g_2(x,\sigma) := \frac{\partial g_1(x,\sigma)}{\partial\sigma} = \frac{1}{\sigma^2}\left[1 + \frac{x}{\sigma}\left\{ 2\frac{h'}{h} + \frac{x}{\sigma}\left(\frac{h'}{h}\right)' \right\} \left(\frac{x}{\sigma}\right) 1_{x\neq 0} \right], \qquad (2.22)$$

and using **A5**, **A6** and **A11**, this term is bounded by

$$\begin{aligned}
&\frac{C_1}{\sqrt{n}\underline{\omega}} \sum_{t=1}^{n} |g_2(\epsilon_t, \sigma_t^*)| \rho^t \left\| \frac{\partial \sigma_t(\theta)}{\partial\theta} \right\| \\
\leq\ & \frac{C_1}{\sqrt{n}\underline{\omega}} \sum_{t=1}^{n} \rho^t \left\{ 1 + 3C_0 \left(1 + \left| \frac{\epsilon_t}{\underline{\omega}} \right|^{\delta} \right) \right\} \sup_{\theta\in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial\theta} \right\| \qquad (2.23)
\end{aligned}$$

where $\sigma_t^* = \sigma_t^*(\theta)$ is between $\widetilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. Using the Cauchy-Schwarz inequality, **A12**, and already given arguments, it can be show that the right-hand side of (2.23) is a.s. equal to $O(1/\sqrt{n})$. It follows that the right-hand side of (2.21) tends to zero, which completes the proof of *iv)*.

Now we establish *v)*. The invertibility of J_* follows from **A9**. Using **A5** and **A11**, we have

$$\begin{aligned}
\left\| \frac{\partial^2 g(\epsilon_t, \sigma_t(\theta))}{\partial\theta\partial\theta'} \right\| &= \left\| g_2(\epsilon_t, \sigma_t(\theta)) \frac{\partial \sigma_t(\theta)}{\partial\theta} \frac{\partial \sigma_t(\theta)}{\partial\theta'} + g_1(\epsilon_t, \sigma_t(\theta)) \frac{\partial^2 \sigma_t(\theta)}{\partial\theta\partial\theta'} \right\| \\
&\leq \left\{ 1 + 3C_0 \left(1 + \left| \frac{\eta_t}{\sigma_*} \right|^\delta \left| \frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)} \right|^\delta \right) \right\} \left(\left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial\theta\partial\theta'} \right\| \right. \\
&\qquad \left. + \left\| \frac{1}{\sigma_t^2(\theta)} \frac{\partial \sigma_t(\theta)}{\partial\theta} \frac{\partial \sigma_t(\theta)}{\partial\theta'} \right\| \right).
\end{aligned}$$

Hence

$$E \sup_{\theta \in V(\theta_0^*)} \left\| \frac{\partial^2 g(\epsilon_t, \sigma_t(\theta))}{\partial\theta\partial\theta'} \right\| < \infty$$

by Hölder's inequality, **A5** and **A12**. The ergodic theorem then implies that

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{\theta \in V(\theta_0^*)} \left\| \frac{\partial^2 Q_n(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0^*)}{\partial\theta\partial\theta'} \right\| \\
\leq\ & E \sup_{\theta \in V(\theta_0^*)} \left\| \frac{\partial^2 g(\epsilon_t, \sigma_t(\theta))}{\partial\theta\partial\theta'} - \frac{\partial^2 g(\epsilon_t, \sigma_t(\theta_0^*))}{\partial\theta\partial\theta'} \right\| \qquad a.s.
\end{aligned}$$

By the dominated convergence theorem, the last expectation tends to zero when the neighborhood $V(\theta_0^*)$ tends to the singleton $\{\theta_0^*\}$. The consistency of $\hat{\theta}_n^*$ thus entails

$$\lim_{n\to\infty} \left| \frac{\partial^2 Q_n(\theta^*)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0^*)}{\partial\theta\partial\theta'} \right| = 0, \quad a.s.$$

Now, note that by **A4**, **A5** and the dominated convergence theorem

$$Eg_1(\eta_0, \sigma_*) = 0, \qquad \text{and thus } Eg_1(\sigma_*^{-1}\eta_0, 1) = 0. \tag{2.24}$$

Moreover, we have

$$g_1\left\{\epsilon_t, \sigma_t(\theta_0^*)\right\} = g_1\left\{\sigma_t(\theta_0)\eta_t, \sigma_*\sigma_t(\theta_0)\right\} = \sigma_t^{-1}(\theta_0^*) g_1(\sigma_*^{-1}\eta_t, 1).$$

It follows that

$$Eg_1\left\{\epsilon_t, \sigma_t(\theta_0^*)\right\} \frac{\partial^2 \sigma_t(\theta_0^*)}{\partial\theta\partial\theta'} = 0.$$

Similarly, in view of (2.22), $g_2(\epsilon_t, \sigma_t(\theta_0^*)) = \sigma_t^{-2}(\theta_0^*) g_2(\sigma_*^{-1}\eta_t, 1)$. We also have $\partial\sigma_t^2(\theta)/\partial\theta = 2\sigma_t(\theta)\partial\sigma_t(\theta)/\partial\theta$. By the ergodic theorem, we then have

$$\lim_{n\to\infty} \frac{\partial^2 Q_n(\theta_0^*)}{\partial\theta\partial\theta'} = \frac{Eg_2(\sigma_*^{-1}\eta_0, 1)}{4} J_*, \quad a.s.$$

and *v)* is established.

To prove *vi)* it suffices to note that, based on arguments used to show *v)*,

$$\sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0^*) = \frac{1}{\sqrt{n}} \sum_{t=1}^n g_1\left(\sigma_*^{-1}\eta_t, 1\right) \frac{1}{2\sigma_t^2(\theta_0^*)} \frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} \tag{2.25}$$

and apply a CLT for square integrable stationary martingale differences (see Billingsley (1961)).

Now, from **A8** and the consistency of $\hat{\theta}_n^*$, a Taylor expansion shows that for n large enough

$$\begin{aligned} 0 &= \sqrt{n}\frac{\partial}{\partial\theta} Q_n(\hat{\theta}_n^*) + \sqrt{n}\frac{\partial}{\partial\theta}\widetilde{Q}_n(\hat{\theta}_n^*) - \sqrt{n}\frac{\partial}{\partial\theta} Q_n(\hat{\theta}_n^*) \\ &= \sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0^*) + \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta^*)\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) \\ &\quad + \sqrt{n}\left(\frac{\partial}{\partial\theta}\widetilde{Q}_n(\hat{\theta}_n^*) - \frac{\partial}{\partial\theta} Q_n(\hat{\theta}_n^*)\right), \end{aligned}$$

where θ^* is between $\hat{\theta}_n^*$ and θ_0^*. Applying *iv)* and *v)* we obtain

$$\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) = \frac{-4}{Eg_2\left(\sigma_*^{-1}\eta_t, 1\right)} J_*^{-1} \sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0^*) + o_P(1). \tag{2.26}$$

and the proof of the asymptotic normality comes from *vi)*.

2.5.2 Proof of Theorem 2.1

Following Koenker (2006), we have

$$\hat{\xi}_{\alpha,n}^* = \arg\min_{z\in\mathbb{R}} \sum_{t=1}^n \rho_\alpha(\hat{\eta}_t^* - z),$$

where $\rho_\alpha(u) = u(\alpha - \mathbf{1}_{u<0})$. Thus

$$\sqrt{n}(\hat{\xi}_{\alpha,n}^* - \xi_\alpha^*) = \arg\min_{z\in\mathbb{R}} O_n(z)$$

where

$$O_n(z) = \sum_{t=1}^n \rho_\alpha\left(\hat{\eta}_t^* - \xi_\alpha^* - \frac{z}{\sqrt{n}}\right) - \sum_{t=1}^n \rho_\alpha(\eta_t^* - \xi_\alpha^*).$$

Let $\eta_t(\theta) = \epsilon_t/\sigma_t(\theta)$. Note that, by **A3** and **A6**, for n large enough

$$\left|\hat{\eta}_t^* - \eta_t(\hat{\theta}_n^*)\right| = \left|\epsilon_t \frac{\sigma_t(\hat{\theta}_n^*) - \widetilde{\sigma}_t(\hat{\theta}_n^*)}{\widetilde{\sigma}_t(\hat{\theta}_n^*)\sigma_t(\hat{\theta}_n^*)}\right| \leq \frac{C_1}{\underline{\omega}}\rho^t \sup_{\theta\in V(\theta_0^*)}\left|\frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)}\right|. \tag{2.27}$$

A Taylor expansion around θ_0^* and **A3**, **A6** yield

$$\hat{\eta}_t^* = \eta_t^* - \eta_t^* D_t'(\hat{\theta}_n^* - \theta_0^*) + r_{t,n}$$

with

$$r_{t,n} = \frac{1}{2}(\hat{\theta}_n^* - \theta_0^*)' \frac{\partial^2\eta_t(\theta^*)}{\partial\theta\partial\theta'}(\hat{\theta}_n^* - \theta_0^*) + \hat{\eta}_t^* - \eta_t(\hat{\theta}_n^*),$$

where $D_t = D_t(\theta_0^*)$ and θ^* is between $\hat{\theta}_n^*$ and θ_0^*.

Using (2.1), we thus have

$$\begin{aligned} O_n(z) &= \sum_{t=1}^n \rho_\alpha\left(\eta_t^* - \xi_\alpha^* - \eta_t^* D_t'(\hat{\theta}_n^* - \theta_0^*) - \frac{z}{\sqrt{n}} + r_{t,n}\right) \\ &\quad -\rho_\alpha(\eta_t^* - \xi_\alpha^*). \end{aligned}$$

Using the identity

$$\rho_\alpha(u-v) - \rho_\alpha(u) = -v(\alpha - \mathbf{1}_{\{u<0\}}) + \int_0^v \left\{\mathbf{1}_{\{u\leq s\}} - \mathbf{1}_{\{u<0\}}\right\} ds$$

for $u \neq 0$ (see Equation (A.3) in Koenker and Xiao, 2006), we then obtain

$$O_n(z) = zX_n + Y_n + Z_n(z) + W_n(z)$$

where

$$\begin{aligned} X_n &= \frac{1}{\sqrt{n}}\sum_{t=1}^n (\mathbf{1}_{\{\eta_t^*<\xi_\alpha^*\}} - \alpha), \quad Y_n = \frac{1}{\sqrt{n}}\sum_{t=1}^n R_{t,n}(\mathbf{1}_{\{\eta_t^*<\xi_\alpha^*\}} - \alpha), \\ Z_n(z) &= \sum_{t=1}^n \int_0^{z/\sqrt{n}} (\mathbf{1}_{\{\eta_t^*\leq\xi_\alpha^*+s\}} - \mathbf{1}_{\{\eta_t^*<\xi_\alpha^*\}})ds, \\ W_n(z) &= \sum_{t=1}^n \int_{z/\sqrt{n}}^{(z+R_{t,n})/\sqrt{n}} (\mathbf{1}_{\{\eta_t^*\leq\xi_\alpha^*+s\}} - \mathbf{1}_{\{\eta_t^*<\xi_\alpha^*\}})ds \end{aligned}$$

with $R_{t,n} = \eta_t^* D_t' \sqrt{n}(\hat{\theta}_n^* - \theta_0^*) - \sqrt{n} r_{t,n}$.

By the change of variable $w = s - z/\sqrt{n}$, we have $W_n(z) = \sum_{i=1}^2 W_n^{(i)}(z)$ in which $W_n^{(i)}(z) = \sum_{t=1}^n W_{n,t}^{(i)}$ and

$$\begin{aligned}
W_{n,t}^{(1)} &= \int_0^{R_{t,n}/\sqrt{n}} (\mathbf{1}_{\{\eta_t^* - \xi_\alpha^* - z/\sqrt{n} \le w\}} - \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* - z/\sqrt{n} < 0\}}) dw, \\
W_{n,t}^{(2)} &= \int_0^{R_{t,n}/\sqrt{n}} \left(\mathbf{1}_{\{\eta_t^* - \xi_\alpha^* - z/\sqrt{n} < 0\}} - \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* < 0\}}\right) dw.
\end{aligned}$$

Note that the integrand in $W_{n,t}^{(2)}$ does not depend on w. Therefore, we have

$$W_{n,t}^{(2)} = \left\{\eta_t^* D_t'(\hat{\theta}_n^* - \theta_0^*) - r_{t,n}\right\} \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}}$$

when $z \ge 0$, and

$$W_{n,t}^{(2)} = -\left\{\eta_t^* D_t'(\hat{\theta}_n^* - \theta_0^*) - r_{t,n}\right\} \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in (z/\sqrt{n}, 0)\}}$$

when $z < 0$.

First consider the case $z \ge 0$. Note that

$$\begin{aligned}
\sum_{t=1}^n W_{n,t}^{(2)} &= \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \eta_t^* \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}} D_t'\right) \sqrt{n}(\hat{\theta}_n^* - \theta_0^*) \\
&\quad + \sqrt{n}(\hat{\theta}_n^* - \theta_0^*)' \frac{1}{2n} \sum_{t=1}^n \frac{\partial^2 \eta_t(\theta^*)}{\partial\theta\partial\theta'} \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}} \sqrt{n}(\hat{\theta}_n^* - \theta_0^*) \\
&\quad + \sum_{t=1}^n \left\{\hat{\eta}_t^* - \eta_t(\hat{\theta}_n^*)\right\} \mathbf{1}_{\{\eta_t^* - \xi_\alpha^* \in [0, z/\sqrt{n})\}}. \qquad (2.28)
\end{aligned}$$

Now note that

$$\frac{\partial^2 \eta_t(\theta)}{\partial\theta\partial\theta'} = -\eta_t^* \frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)} \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial\theta\partial\theta'} + 2\eta_t^* \frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)} \frac{1}{\sigma_t^2(\theta)} \frac{\partial \sigma_t(\theta)}{\partial\theta} \frac{\partial \sigma_t(\theta)}{\partial\theta'}.$$

In view of **A5** and the last part of **A12**, for $\theta \in V(\theta_0^*)$, $\eta_t^* \frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)}$ admits a moment larger than 2. The first part of **A12** and the Cauchy-Schwartz inequality then entail that

$$E \sup_{\theta \in V(\theta_0^*)} \left\| \frac{\partial^2 \eta_t(\theta)}{\partial\theta\partial\theta'} \right\|^{1+\nu} < \infty$$

for some $\nu > 0$. By Hölder's inequality, for $\theta^* \in V(\theta_0^*)$,

$$\begin{aligned} & \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \eta_t(\theta^*)}{\partial\theta\partial\theta'} \mathbf{1}_{\{\eta_t^*-\xi_\alpha^* \in [0,z/\sqrt{n})\}} \right\| \\ \leq \quad & \left\{ \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in V(\theta_0^*)} \left\| \frac{\partial^2 \eta_t(\theta^*)}{\partial\theta\partial\theta'} \right\| \right\}^{1/(1+\nu)} \left\{ \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{\eta_t^*-\xi_\alpha^* \in [0,z/\sqrt{n})\}} \right\}^{\nu/(1+\nu)} \end{aligned}$$

The term $\sum_{t=1}^{n} \mathbf{1}_{\{\eta_t^*-\xi_\alpha^* \in [0,z/\sqrt{n})\}} = O_P(\sqrt{n})$ because its expectation is $O(\sqrt{n})$ and its variance is $O(\sqrt{n})$. It follows that the second term on the right-hand side of (2.28) tends to zero in probability. By the same arguments and (2.27), we show that the third term has the same behavior.

Now, noting that $\xi_\alpha^* f^*(\xi_\alpha^*) = \xi_\alpha f(\xi_\alpha)$ when f^* is the density of $\eta_1^* = \eta_1/\sigma_*$, we have

$$\begin{aligned} E(\eta_t^* \mathbf{1}_{\{\eta_t^*-\xi_\alpha^* \in [0,z/\sqrt{n})\}} &= \int_0^{z/\sqrt{n}} (x+\xi_\alpha^*) f^*(x+\xi_\alpha^*) dx \\ &= \xi_\alpha f(\xi_\alpha) \frac{z}{\sqrt{n}} + o(1/\sqrt{n}). \end{aligned}$$

Thus, in view of the independence of η_t^* and D_t, we have

$$E\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_t^* \mathbf{1}_{\{\eta_t^*-\xi_\alpha^* \in [0,z/\sqrt{n})\}} D_t' \right) = z\xi_\alpha f(\xi_\alpha) \Omega_*' + o(1).$$

By similar computations we find

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_t^* \mathbf{1}_{\{\eta_t^*-\xi_\alpha^* \in [0,z/\sqrt{n})\}} D_t' \right) = o(1).$$

It follows that

$$\sum_{t=1}^{n} W_{n,t}^{(2)} = z\xi_\alpha f(\xi_\alpha) \Omega_*' \sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o(1) \quad a.s.$$

The same equality holds for $z \leq 0$.

We now denote by $E_{t-1}X$ the expectation of some variable X conditional on $\{\hat{\theta}_n^* - \theta_0^*, (\eta_u : u < t)\}$. We have, by the change of variable $w = \eta_t^* v$,

$$\begin{aligned} E_{t-1} W_{n,t}^{(1)} &= \int_0^{D_t'(\hat{\theta}_n^*-\theta_0^*)+o(n^{-1/2})} E_{t-1}(\eta_t^* \mathbf{1}^*_{\{\eta_t^* \in (\xi_\alpha^*+z/\sqrt{n}, (\xi_\alpha^*+z/\sqrt{n})(1-v)^{-1})\}}) dv \\ &= \frac{(\xi_\alpha^*)^2}{2} f_{n,t}^*(\xi_\alpha^*)(\hat{\theta}_n^* - \theta_0^*)' D_t D_t' (\hat{\theta}_n^* - \theta_0^*) + o(n^{-1}) \quad a.s. \end{aligned}$$

where $f^*_{n,t}$ denotes the density of η^*_t conditional on $\{\hat{\theta}^*_n - \theta^*_0, (\eta_u : u < t)\}$ and $o(n^{-1})$ is a function of $(\hat{\theta}^*_n - \theta^*_0)$ and the past values of η^*_t. By the arguments used for $X^{(2)}_{n,t}$ it can therefore be shown that $W^{(1)}_n(z)$ converges in distribution to a random variable which does not depend on z. Note also that Y_n can be subtracted from the objective function $O_n(z)$ because it does not depend on z. Moreover $Z_n(z) \to \frac{z^2}{2} f^*(\xi^*_\alpha)$ in probability as $n \to \infty$. Finally,

$$\tilde{O}_n(z) := O_n(z) - Y_n = \frac{z^2}{2} f^*(\xi^*_\alpha) + z\{X_n + \xi^*_\alpha f(\xi_\alpha)\Omega'_*\sqrt{n}(\hat{\theta}^*_n - \theta^*_0)\} + O_P(1).$$

Since the process $\tilde{O}_n(\cdot)$ has convex sample paths, the convexity Lemmas of Knight (1989) and Pollard (1991) show that $\tilde{O}_n$ converges weakly to some convex process. By Lemma 2.2 in Davis et al. (1992), we can conclude that

$$\begin{aligned}\sqrt{n}(\hat{\xi}^*_{\alpha,n} - \xi^*_\alpha) &= -\xi^*_\alpha\Omega'_*\sqrt{n}(\hat{\theta}^*_n - \theta^*_0) \\ &\quad - \frac{1}{f^*(\xi^*_\alpha)}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbf{1}_{\{\eta^*_t < \xi^*_\alpha\}} - \alpha) + o_P(1).\end{aligned}$$

In view of (2.25) and (2.26), we have

$$\sqrt{n}(\hat{\theta}^*_n - \theta^*_0) = \frac{-4}{Eg_2(\eta^*_0, 1)} J^{-1}_* \frac{1}{\sqrt{n}}\sum_{t=1}^n g_1(\eta^*_t, 1) D_t(\theta^*_0) + o_P(1).$$

By the CLT for martingale differences, we get

$$S_n := \frac{1}{\sqrt{n}}\sum_{t=1}^n \begin{pmatrix} g_1(\eta^*_t, 1)D_t(\theta^*_0) \\ \mathbf{1}_{\{\eta^*_t < \xi^*_\alpha\}} - \alpha \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \begin{pmatrix} \frac{Eg_1^2(\eta^*_1,1)}{4} J_* & c_\alpha \Omega_* \\ c_\alpha\Omega'_* & \alpha(1-\alpha) \end{pmatrix}\right\}.$$

The result follows from

$$\sqrt{n}\begin{pmatrix} \hat{\theta}^*_n - \theta^*_0 \\ \hat{\xi}^*_{\alpha,n} - \xi^*_\alpha \end{pmatrix} = \begin{pmatrix} \frac{-4}{Eg_2(\eta^*_0,1)} J^{-1}_* & 0_m \\ \frac{4\xi^*_\alpha}{Eg_2(\eta^*_0,1)}\Omega'_* J^{-1}_* & \frac{-1}{f^*(\xi^*_\alpha)} \end{pmatrix} S_n,$$

using the relation $\Omega'_* J^{-1}_* \Omega_* = 1/4$ by Remark 3.1 in FZ.

□

2.5.3 Proof of Corollary 2.2

In view of (2.5), when h is replaced by h_s, then $\hat{\theta}^*_n$ is replaced by $\hat{\theta}^{(s)}_n$ such that $\hat{\theta}^*_n = H(\hat{\theta}^{(s)}_n, s)$. It is then clear that $\tilde{\sigma}_t(\hat{\theta}^*_n)$ and η^*_t are replaced by respectively $\tilde{\sigma}_t(\hat{\theta}^*_{n,s}) = s^{-1}\tilde{\sigma}_t(\hat{\theta}^*_n)$ and $s\eta^*_t$, and thus the VaR estimator is unchanged.

References

Artzner, P., Delbaen, F., Eber, J-M. and D. Heath (1999) Coherent measures of risk. *Mathematical Finance* 9, 203–228.

Bardet, J-M. and O. Wintenberger (2009) Asymptotic normality of the Quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics* 37, 2730–2759.

Bassett, G-W. and R-W. Koenker (1986) Strong Consistency of Regression Quantiles and Related Empirical Processes. *Econometric Theory* 2, 191–201.

Berkes, I. and L. Horvâth (2004) The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* 32, 633–655.

Berkes, I., Horvâth, L. and P. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

Billingsley, P. (1961) *Statistical inference for Markov processes.* Chicago: University of Chicago Press.

Billingsley, P. (1995) *Probability and Measure.* Third edition John Wiley and Sons.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Christoffersen, P.F. (2003) *Elements of financial risk management.* Academic Press, London.

Davis, R.A., Knight, K. and J. Liu (1992) M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications* 40, 145–180.

Ding, Z., Granger C. and R.F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987–1008.

Escanciano, J.C. and J. Olmo (2010) Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics* 28, 36–51.

Fan, J., Qi, L. and D. Xiu (2014) Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics* 32, 178–191.

Francq, C., Lepage, G. and J-M. Zakoïan (2011) Two-stage non Gaussian QML estimation of GARCH Models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics* 165, 246–257.

Francq, C. and J-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. and J-M. Zakoïan (2012) Risk-parameter estimation in volatility models. MPRA Preprint No. 41713.

Francq, C. and J.M. Zakoïan (2013) Optimal predictions of powers of conditionally heteroskedastic processes. *Journal of the Royal Statistical Society - Series B* 75, 345–367.

Glosten, L., Jagannathan, R. and D. Runkle (1993) Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

Hamadeh, T. and J-M. Zakoïan (2011)Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH Processes. *Journal of Statistical Planning and Inference* 141,488–507.

Knight, K. (1998) Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics* 26, 755–770.

Koenker, R. (2005) *Quantile Regression.* Cambridge: Cambridge University Press.

Koenker, R. and Z. Xiao (2006) Quantile autoregression. *Journal of the American Statistical Association* 101, 980–990.

Kuester, K., Mittnik, S. and M.S. Paolella (2006) Value-at-Risk predictions: A comparison of alternative strategies. *Journal of Financial Econometrics* 4, 53–89.

Lee, S.W. and B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* 10, 29–52.

Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.

McNeil, A.J., Frey, R. and P. Embrechts (2005) *Quantitative Risk Management.* Princeton University Press.

Mikosch, T. and D. Straumann (2006) Stable limits of martingale transforms with application to the estimation of GARCH parameters. *The Annals of Statistics* 34, 493–522.

Pollard, D. (1991) Asymptotics for Least Absolute Deviation Regression Estimators *Econometric Theory* 7, 186–199.

Straumann, D. and T. Mikosch (2006) Quasi-maximum likelihood estimation in conditionally heteroscedastic Time Series: a stochastic recurrence equations approach. *The Annals of Statistics* 5, 2449–2495.

Wang, S. (2000) A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance* 67, 15–36.

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

Wirch, J. L. and M.R. Hardy (1999) A Synthesis of Risk Measures for Capital Adequacy. *Insurance: Mathematics and Economics* 25, 337–347.

Xiao, Z. and R. Koenker (2009) Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association* 104, 1696–1712.

Xiao, Z. and C. Wan (2010) A robust estimator of conditional volatility. Unpublished document.

Zakoïan J-M. (1994) Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control* 18, 931–955.

Zhu, K. and S. Ling (2011) Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *The Annals of Statistics* 32, 2131–2163.

Chapitre 3

Testing the efficiency of a generalized QMLE for conditional VaR estimation

Abstract. This chapter studies the efficiency of conditional value at risk (VaR) parameter estimation. We consider two methods namely the gaussian quasi maximum likelihood and the generalized quasi maximum likelihood (gQML) based on a double generalized Gamma instrumental density (dgG). We seek to test the optimality of the VaR parameter estimator based on the Gaussian distribution against the VaR parameter estimator based on a dgG. Numerical illustrations are provided through simulation experiments and applications to financial stock indexes, exchange rates and commodities.

KEYWORDS. Conditional VaR, Double generalised Gamma distribution, GARCH models, Generalized Quasi Maximum Likelihood Estimation

3.1 Introduction

Several researchers studied the asymptotic properties of estimators of commonly used GARCH models (2.4) such as the maximum likelihood estimator(MLE), Gaussian quasi maximum likelihood estimator and Non Gaussian QMLE (see, Lee and Hansen (1994), Hall and Yao (2003), Berkes and Horvàth (2004), Francq and Zakoïan (2004), Francq and Zakoïan (2013) and the references therein).

Recently, Francq et al. (2011) studied the consistency and the asymptotic normality (CAN) of the generalized QMLE (gQMLE) based on a particular class of instrumental density h. They have shown that the gQMLE is more efficient than the standard QMLE. Their two-step approach for inferring GARCH(p,q) model consists on estimating a reparametrized model under the identifiability condition $E|\eta_t|^r = 1$. The CAN properties of their estimator depend on the power r for a given class of instrumental densities of the form $h_r(x) = \frac{r^{1-1/r}}{2\Gamma(1/r)}exp(-\frac{1}{r}|x|^r)$, given that $r > 0$. For optimal values of r they tested whether the gQMLE has better accuracy than the Gaussian QMLE within the class of estimators. For $r = 2$, their gQMLE coincide with the standard Gaussian case, otherwise for optimal values of $r \neq 2$ they found that the gQMLE is more accurate .

Francq and Zakoïan (2012) studied CAN properties of the gQMLE for risk parameter (called VaR parameter) estimation under mild regularity assumptions. They proposed a one step approach based on the gQMLE for conditional VaR parameter estimation and have compared it to a two step approach based on the Gaussian QMLE.

In chapter 2 we considered a two-step approach based on the gQMLE for conditional Value at Risk (VaR) estimation. As detailed in chapter 2 given a class $\mathbb{H}$ of instrumental densities, where $h \in \mathbb{H}$ such that $h \neq P_\eta$, and P_η is the distribution of the residuals, the gQMLE converges to a "pseudo-true" value θ_0^*. Noting that, any reasonable GARCH type model of the form (2.4), is stable by scaling which is conform with the following assumption:

A1: There exists a function H such that for any $\theta \in \Theta$, for any $K > 0$, and any sequence (x_i)

$$K\sigma(x_1, x_2, \ldots; \theta) = \sigma(x_1, x_2, \ldots; H(\theta, K)).$$

Under **A1**, considering a GARCH(p,q) model without a well defined volatility parameter θ_0 and an unknown distribution of the residuals P_η, θ_0^* can be written as $\theta_0^* = H(\theta_0, \sigma_*)$. Hence, it depends on θ_0 and a scale parameter $\sigma_* > 0$ depending on h and P_η.

Based on the results of chapter 2, without assuming that $E\eta_t^2 = 1$ and the unique maximum $\sigma_* = 1$ the estimation of the conditional VaR returns to estimate the VaR parameter $\theta_{0,\alpha} = H(\theta_0, -\xi_\alpha)$, where ξ_α denotes the theoretical α-quantile of η_t. We have shown that this method leads to a consistent estimation of the conditional VaR even if $h \neq P_\eta$. The asymptotic distribution of the two-step estimator is also derived. In order to obtain the asymptotic distribution of the two-step VaR parameter estimator, in an intermediate step, a joint limiting distribution of the volatility parameter θ_0 and the α-quantile of the innovations distribution has been derived yielding an explicit form of the variance matrix Σ^h which depends on h (see Theorem 2.1 in chapter 2). Then, the asymptotic variance of the two-step VaR estimator is achieved. In chapter 2 we prove that the function h that minimizes the asymptotic variance is the optimal instrumental density h_{opt}. It is proved that h_{opt} depends neither on the GARCH parameter θ_0 nor on the risk level α, but only on the simple characteristics of P_η. The ranking of two instrumental densities h_1 and h_2 is related to the asymptotic variance, where h_1 can be better than h_2 ($h_1 \succ h_2$) if the difference between the asymptotic variances is definite positive. It remains problematic to know if the optimal distribution performs better than the Gaussian instrumental distribution $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ or not (*i.e.* the optimal distribution h_1 that gives the minimum variance corresponds to ϕ or $h_1 \succ \phi$).

We want to give a formal test of whether the standard two-step approach method in which $h = \phi$ is more accurate than the gQMLE. We define a hypothesis which states that the Gaussian QMLE is more optimal than the gQMLE.

If this hypothesis will be rejected, thereby the adaptive approach based on the optimal instrumental density h is better. Hence, we consider as instrumental density the double generalized Gamma distribution with positive parameters b, p and d denoted by $h_{dgG} = \Gamma(b, p, d)$. It can be considered as a large class of instrumental densities $\mathbb{H}_{dgG}$ including the Gaussian distribution for $b = \frac{1}{\sqrt{2}}$, $p = 1$ and $d = 2$. Under some assumptions and by using h_{dgG} a consistent gQML estimator of the volatility parameter θ_0 of the model (2.4) is obtained. Consequently, we can derive a consistent conditional VaR estimator based on the two-step approach. The asymptotic variance of this estimator, as detailed in chapter 2, does not depend on θ_0 but only on the parameter d of h_{dgG} and some moments of η_t. Thereby, to find the most useful $h \in \mathbb{H}_{dgG}$, we have to find the optimal d, denoted by d_{opt}, which is the parameter that minimizes the asymptotic variance of the gQMLE. Then, we notice the importance of defining a consistent estimator of d_{opt} and derive its asymptotic variance. Thereafter, formally our hypothesis can be written as $H_0 : d_{opt} = 2$.

The reminder of this chapter is as follows: Section 3.2 provides the asymptotic confidence interval of the VaR. Section 3.3 studies, within a class of instrumental densities called $dgG(b, p, d)$, the asymptotic distribution of the parameter d that minimizes the asymptotic variance of the VaR parameter estimator. The theoretical illustrations, simulation results and real case applications are displayed respectively in Section 3.4, Section 3.5 and Section 3.6. A brief summary of our results is given in Section 3.7. The proofs are displayed in Section 3.8.

3.2 A gQML-two-step-approach for conditional VaR estimation

By using Lemma 2.1, Theorem 2.1 and Corollary 2.1 in chapter 2 we establish the following Lemma in order to give an explicit form of the Confidence interval for the conditional VaR.

Lemma 3.1 [Confidence interval of the conditional VaR] *Under the assumption* **C**, $\hat{\theta}^*_{n,\alpha}$ *converges a.s. to* $\theta_{0,\alpha}$ *for* $n \to \infty$ *and*

$$\sqrt{n}\left(\hat{\theta}^*_{n,\alpha} - \theta_{0,\alpha}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, G_h \Sigma^h {G_h}'\right).$$

By using the delta method, we obtain a confidence interval of the $VaR_t(\alpha)$ *at the statistical level* $\underline{\alpha}$. *We note that there is difference between statistical risk* $\underline{\alpha}$ *and financial risk* α. *Let* $\hat{\Xi}_\alpha$ *be a consistent estimator of the asymptotic variance* $G_h\Sigma^h {G_h}'$, *and* $\mu_{\underline{\alpha}}$ *is the* $(1 - \underline{\alpha}/2)$ *quantile of the normal distribution, we obtain*

$$S_n := \left[V\hat{a}R_t(\alpha) \pm \frac{\mu_{\underline{\alpha}}}{\sqrt{n}} \sqrt{\frac{\partial \hat{\sigma}_t(\hat{\theta}_{n,\alpha})}{\partial \theta'} \hat{\Xi}_\alpha \frac{\partial \hat{\sigma}_t(\hat{\theta}_{n,\alpha})}{\partial \theta}} \right]. \tag{3.1}$$

Remark 3.1 *In the asymptotic variance given by Equation (2.2), only* τ_h *given by Equation (2.9) depends on* h. *Thus, to obtain the optimal* τ_h, *we should find the optimal* h *that minimizes this term. With reference to Remark 2.3 in chapter 2 it is shown that the optimality of* h *depends neither on the volatility model nor on the risk level* $\underline{\alpha}$, *however it depends on some characteristics of* P_η.

Example 3.1 (Gaussian distribution) *Let for any* $r > 0$, $\hat{\mu}_r = \frac{1}{n}\sum_{t=1}^n |\hat{\eta}_t|^r$ *with* $\hat{\eta}_t = \frac{\epsilon_t}{\sigma_t(\hat{\theta}_n)}$.

For $h = \phi$, $\hat{\theta}_n$ *is the Gaussian QMLE. Given Equation (2.9),* $\sigma_* = 1$, $E\eta_t^2 = 1$ *and* $\kappa_4 := E\eta_t^4 < \infty$ *we have*

$$\tau_\phi = \left(\frac{E|\eta_t|^4}{(E|\eta_t|^2)^2} - 1\right) = \kappa_4 - 1, \quad \text{and} \quad \hat{\tau}_\phi = \left(\frac{\hat{\mu}_2^2}{(\hat{\mu}_2)^2} - 1\right).$$

We obtain the same results as Theorem 4.2 in Francq and Zakoïan (2012) and the asymptotic variance follows

$$G_\phi \Sigma^\phi {G_\phi}' = (\kappa_4 - 1)A(J^{-1} - \Psi)A + \frac{4\xi_\alpha^2 \alpha(1-\alpha)}{f^2(\xi_\alpha)}\Psi.$$

Corollary 3.1 *The asymptotic efficency of the conditional VaR estimator $\hat{\theta}^*_{n,\alpha}$ based on the gQMLE can be more accurate than the estimator $\hat{\theta}^*_{n,\alpha}$ based on Gaussian QMLE if and only if (iff),*

$$\frac{\tau_h}{\tau_\phi} < 1.$$

3.3 Efficiency gain of gQMLE based on Double Generalized Gamma distribution over Gaussian QMLE

In Fan et al. (2013) and Francq at al. (2011) among others, the main problem that they discussed is not only to find a consistent estimator of volatility parameters using non Gaussian QMLE but also to evaluate the accuracy of this estimator. Under weak moment conditions, Fan et al. (2013) have studied the efficiency of their three-step approach for GARCH parameters estimation using a non Gaussian QMLE. They have shown that their proposed method based on different likelihood using heavy tailed distributions achieve better precision than the standard Gaussian QMLE. Francq et al. (2011) have estimated a reparametrized form of model (2.4) using the gQMLE for a given class $h_r(x) = \frac{r^{1-1/r}}{2\Gamma(1/r)} exp(-\frac{1}{r}|x|^r)$, for $r > 0$ .This class of distributions is namely the generalized Gaussian distribution. Based on this class of distributions and given Equation (2.9), $\tau_{h_r} = \frac{4}{r^2}\left(\frac{E|\eta_1|^{2r}}{(E|\eta_1|^r)^2} - 1\right) = \frac{4}{r^2}\left(\frac{\mu_{2r}}{\mu_r^2} - 1\right)$. Francq et al. (2011) used a test in order to determine the more accurate QMLE. Formally, they tested if the optimal power r which minimizes τ_{h_r} coincide with the Gaussian case for $r = 2$ or not. For a given significance level α_0, their null hypothesis states that the Gaussian QMLE is more efficient.

In our framework, we want to test the efficiency of the gQMLE in estimating the conditional VaR at a given risk level $\underline{\alpha}$. Under Remark 3.1 the optimal $h \in \mathbb{H}$ is the distribution that minimizes τ_h. For this distribution we will obtain the optimal τ_h denoted τ_{hopt}, by

$$\tau_{hopt} = \min_{h\in\mathbb{H}} \frac{4Eg_1^2(\sigma_*^{-1}\eta_0, 1)}{\left\{Eg_2(\sigma_*^{-1}\eta_0, 1)\right\}^2}. \tag{3.2}$$

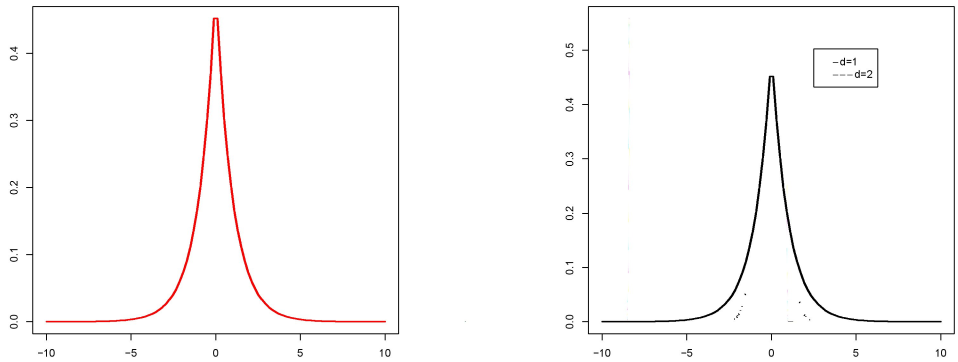
Let h be a dgG distribution $\Gamma(b, p, d)$ with parameters $b > 0$, $p > 0$ and $d > 0$ defined by the density

$$h(x) = h_{dgG}(x) = \frac{db^p}{2\Gamma(\frac{p}{d})}|x|^{p-1}e^{-|bx|^d}.$$

This distribution can be considered as a large class of instrumental densities $\mathbb{H}_{dgG}$. It contains in particular the Laplace distribution

$$h_l(x) = e^{-|x|}/2,$$

with a null location parameter and a scale parameter equal to 1 that corresponds to the $\Gamma(1,1,1)$ as illustrated in Figure 3.1.



(a) (b)

Figure 3.1 – *The Laplace probability density function plot (left panel) and the double generalized Gamma distribution plot with parameters $b = 1$, $p = 1$, $d = 1$ and $d = 2$ (right panel).*

It is clear from Figure 3.1 that the dGg with parameters $b = 1$, $p = 1$, $d = 1$ has the same shape as the Laplace distribution.

The generalized error distribution of the shape parameter $\kappa > 0$, denoted by GED(κ) coincides with $\Gamma((\frac{1}{2})^{\frac{1}{\kappa}}, 1, \kappa)$ and it is given by

$$h_{GED_\kappa}(x) = \frac{\kappa}{\Gamma(1/\kappa)2^{1+1/\kappa}} e^{-\frac{|x|^\kappa}{2}},$$

the Gaussian distribution is the $\Gamma(\frac{1}{\sqrt{2}}, 1, 2)$ and it has the following form

$$h(x) = \phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}.$$

The double Weibull distribution, with shape parameter $k > 0$ and scale parameter $\lambda > 0$ denoted by Dweib, is defined as follows

$$h_{Dweib}(x) = \frac{1}{2}\frac{k}{\lambda}|\frac{x}{\lambda}|^{k-1}e^{-|\frac{x}{\lambda}|^k}.$$

It coincides with $\Gamma(\frac{1}{\lambda}, k, k)$. In addition to the above cited distributions, the dgG contains also the Rayleigh and Maxwell distributions. For $h \in \mathbb{H}_{dgG}$, and $x \neq 0$, we have

$$h'_{dgG}(x) = \frac{p-1-d|bx|^d}{|x|} h_{dgG}(x),$$

$$x\frac{h'_{dgG}}{h_{dgG}}(x) = p-1-d|bx|^d.$$

Also we have $\sigma_* = \left(\frac{db^d}{p}E|\eta_1|^d\right)^{1/d}$. Thus

$$g_1\left(\frac{\eta_1}{\sigma_*},1\right) = p\left(\frac{|\eta_1|^d}{E\,|\eta_1|^d}-1\right), \quad g_2\left(\frac{\eta_1}{\sigma_*},1\right) = p\left(1-(d+1)\frac{|\eta_1|^d}{E\,|\eta_1|^d}\right).$$

Then, we have

$$\tau_{dgG} = \tau(d) = \frac{4}{d^2}\left(\frac{E\,|\eta_1|^{2d}}{\left(E\,|\eta_1|^d\right)^2}-1\right). \tag{3.3}$$

Considering Equation (3.2), τ_{dgG} that corresponds to the optimal dgG instrumental density denoted by τ^*_{dgG} is defined by

$$\tau^*_{dgG} = \min_{h\in\mathbb{H}_{dgG}} \frac{4}{d^2}\left(\frac{E\,|\eta_1|^{2d}}{\left(E\,|\eta_1|^d\right)^2}-1\right).$$

Let for any $d>0$, $\mu_d = E|\eta_t|^d$ then

$$\tau^*_{dgG} = \min_{h\in\mathbb{H}_{dgG}} \frac{4}{d^2}\left(\frac{\mu_{2d}}{{\mu_d}^2}-1\right).$$

The optimization problem depends only on the parameter d and on some moments of the innovations process. If the theoretical moments are replaced by emperical ones and if d does not belong to bounded space, the optimisation problem solution diverges (see Francq et al. 2011, Lemma 3.1). To avoid this problem we have to assume the following

B1: There exists a unique $d_{opt} > 0$ such that $d_{opt} = \arg\min_{d_{opt}\in\mathbb{D}} \tau_{dgG}$,

$$d_{opt} = \arg\min_{d_{opt}\in\mathbb{D}} \frac{4}{d^2}\left(\frac{E\,|\eta_1|^{2d}}{\left(E\,|\eta_1|^d\right)^2}-1\right),$$

where $\mathbb{D}$ is the compact parameter space of the form $\mathbb{D} = [\underline{d}, \bar{d}] \in (0, d_{max})$ with $d_{max} = \sup\{d \in \mathbb{R}; \mu_{2d} < \infty\}$.

For $u >$, $v \geq 0$, we note

$$m(u,v) = E\left\{|\eta_t|^u (log|\eta_t|)^v\right\}, \tag{3.4}$$

when the expectation exists.

B2: The support Ω_η of η_t contains at least five values and

$$E|\eta_t|^{4d_{opt}}(log|\eta_t|)^2 \leq \infty.$$

Remark 3.2 *Using* **B2**, *when, $u > 0$ and $v \geq 0$, we have,*

$$0^u(log|0|)^v = 0. \tag{3.5}$$

Let,

$$m_n(u,v,\theta) = \frac{1}{n}\sum_{t=1}^{n} |\frac{\epsilon_t}{\sigma_t(\theta)}|^u (log|\frac{\epsilon_t}{\sigma_t(\theta)}|)^v, \tag{3.6}$$

and

$$\tau_n(d,\theta) = \frac{4}{d^2}\left(\frac{m_n(2d,0,\theta)}{m_n(d,0,\theta)^2} - 1\right). \tag{3.7}$$

Replacing η_t by $\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}$ or the Gaussian residual process $\hat{\eta}_t = \frac{\epsilon_t}{\tilde{\sigma}_t(\hat{\theta}_{n,\phi})}$ in Equation (3.6) we obtain respectively:

$$\hat{m}_n(u,v) = \frac{1}{n}\sum_{t=1}^{n} |\hat{\eta}_t|^u (log|\hat{\eta}_t|)^v,$$

$$\widetilde{m}_n(u,v,\theta) = \frac{1}{n}\sum_{t=1}^{n} |\frac{\epsilon_t}{\widetilde{\sigma}_t(\theta)}|^u (log|\frac{\epsilon_t}{\widetilde{\sigma}_t(\theta)}|)^v.$$

Then, we have

$$\widetilde{\tau}_n(d,\theta) = \frac{4}{d^2}\left(\frac{\widetilde{m}_n(2d,0,\theta)}{\widetilde{m}_n(d,0,\theta)^2} - 1\right).$$

For $n \to \infty$, from Equation (3.6) we obtain

$$m_\infty(u,v,\theta) = E|\frac{\epsilon_t}{\sigma_t(\theta)}|^u (log|\frac{\epsilon_t}{\sigma_t(\theta)}|)^v.$$

B3: d_{opt} belongs to the interior of $\mathbb{D}$.

Let $\hat{\eta}_t = \epsilon_t/\tilde{\sigma}_t(\hat{\theta}_n)$, $t = 1, \ldots, n$, obtained from the Gaussian QMLE

$$\hat{d}_n = \arg\min_{d \in \mathbb{D}} \frac{1}{d^2}\left(\frac{\hat{\mu}_{2d}}{\hat{\mu}_d^2} - 1\right), \quad \hat{\mu}_r = \frac{1}{n}\sum_{t=1}^{n} |\hat{\eta}_t|^r$$

For any fixed n

$$\frac{1}{d^2}\left(\frac{\hat{\mu}_{2d}}{\hat{\mu}_d^2} - 1\right) \to 0, \qquad \text{as } d \to \infty.$$

For this reason, we assume in **B1** that d belongs to a bounded interval $\mathbb{D}$.
The next result show the consistency and asymptotic properties of $\hat{d}_n$.

Theorem 3.1 *Under assumptions* ***B1, B2, B3*** *and* ***C****, we have*

$$\hat{d}_n \quad \to \quad d_{opt} \quad a.s$$

and

$$\sqrt{n}(\hat{d}_n - d_{opt}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Upsilon_{d_{opt}}), \tag{3.8}$$

when using Equations (3.34) and (3.35), Υ_{hopt} is of the form

$$\Upsilon_{d_{opt}} = \frac{\zeta_{d_{opt}}}{\left\{\frac{\partial^2 \tau_\infty}{\partial d^2}(d_{opt}, \theta_0)\right\}^2}. \tag{3.9}$$

Corollary 3.2 *Under the hypothesis of Theorem 3.1, for a given confidence level* $1 - \underline{\alpha}$ *the null hypothesis* $d_{opt} = 2$ *is rejected if*

$$\left\{\frac{n}{\hat{\Upsilon}_{d_n}}(\hat{d}_n - 2)^2 > \left\{\Phi^{-1}(1 - \underline{\alpha}/2)\right\}^2\right\}, \tag{3.10}$$

where $\hat{\Upsilon}_{d_n} = \frac{\hat{\zeta}_{d_{opt}}}{\left\{\frac{\partial^2 \hat{\tau}_\infty}{\partial d^2}(d_{opt}, \theta_0)\right\}^2}$ *and* Φ *is the cumulative distribution function of the standard normal distribution.*

The following Corollary gives an explicit form of the asymptotic confidence interval of d_{opt}.

Corollary 3.3 *Let* $\hat{d}_n$ *be a consistent estimator of the unknown optimal parameter* d_{opt} *and* $\hat{\Upsilon}_{d_n}$ *be a consistent estimator of the asymptotic variance* $\Upsilon_{d_{opt}}$*. The set*

$$I_n := \left[\hat{d}_n - \Phi^{-1}_{1-\underline{\alpha}/2}\sqrt{\frac{\hat{\Upsilon}_{d_n}}{n}}; \hat{d}_n + \Phi^{-1}_{1-\underline{\alpha}/2}\sqrt{\frac{\hat{\Upsilon}_{d_n}}{n}}\right], \tag{3.11}$$

is the confidence interval of d_{opt} *for an asymptotic confidence level* $\underline{\alpha} \in]0, 1[$, i.e.

$$\lim_{n\to\infty} P[d_{opt} \in I_n] = 1 - \underline{\alpha}.$$

3.4 Theoretical illustrations

In this section we illustrate the convergence of $\hat{d}_n$ to d_{opt} and we compare the performance of the gQMLE with the standard Gaussian QMLE.

Example 3.2 **(P_η is the Gaussian distribution)** *We assume that the true density of η_t is the standard gaussian density and the instrumental distribution is the GED(κ) which can be considered as the dgG$((\frac{1}{2})^{\frac{1}{\kappa}}, 1, \kappa)$. Note that for the gaussian density, $\tau_\phi = \mu_4 - 1 = 2$, and it behaves similarly to the GED distribution for $\kappa = 2$. Thus, we should obtain the minimum of τ_{GED} at $\kappa = 2$ which is conform with Figure 3.2. This example shows that the Gaussian QMLE is not outperformed for $\kappa \neq 2$.*

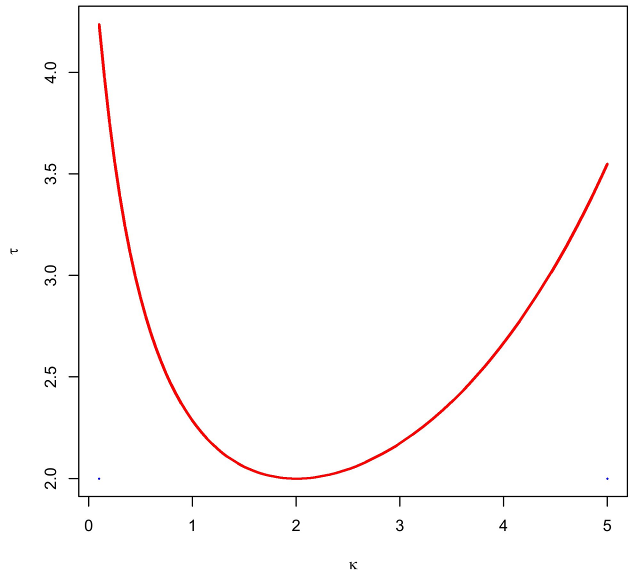


Figure 3.2 – *Optimal power κ_{opt} when η_t has a gaussian distribution, where τ_ϕ is blue and τ_{GED} is red.*

Example 3.3 **(P_η is the GED distribution with $\kappa_0 = 4$)** *Now consider the case when the true density is the GED(4). Always we have obtained the convergence of $\hat{\kappa}$ to κ_0, where the lowest value of τ_{GED} is obtained for $\hat{\kappa} = 4$. It is shown in Figure 3.3 that τ_{GED} has lower values than τ_ϕ for different values of κ.*

Example 3.4 **(The distribution of P_η is the dgG (b, p, d_{opt}))** *It is important to note that the moment of the dgG depends on p and d. We consider here the case*

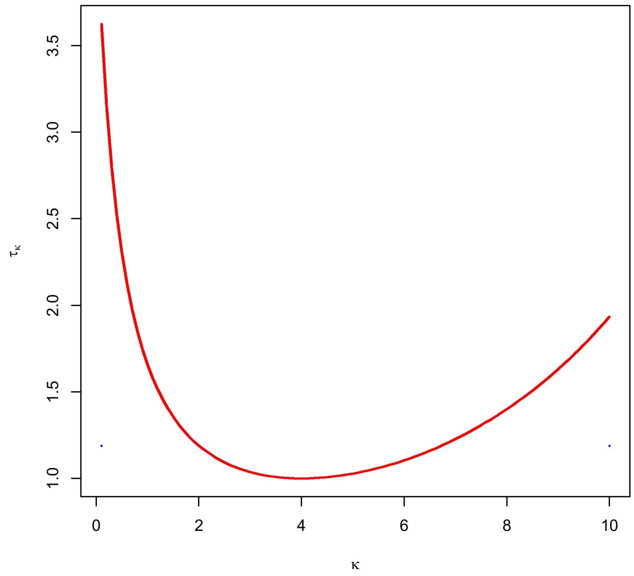


Figure 3.3 – *Optimal power κ_{opt} when the distribution of η_t is the GED(4), where τ_ϕ is blue and τ_{GED} is red.*

where $b = \frac{1}{\sqrt{2}}, p = 1 and \quad d_{opt} = 3$. Figure 3.4 shows that $\hat{\kappa} = \underset{\kappa \in \mathbb{D}}{argmin} \hat{\tau}_{GED} = 3$. There is a gain of efficiency of the gQMLE against the standard Gaussian QMLE for some values of κ near to the optimal value.

3.5 Simulation study

In this section we investigate the efficiency gain of the gQML in the case of misspecification against the standard Gaussian QML and we test here the performance of the Gaussian QML. We consider two different cases of simulation, the first one is when d_{opt} is known and the second when d_{opt} is unknown.

3.5.1 Test of optimality of the Gaussian QMLE when $d_{opt} = 2$

Consider $d_{opt} = 2$, we will follow the four steps cited below in order to test the optimality of the Gaussian QMLE when $\eta_t \sim dgG(\frac{1}{\sqrt{2}}, 1, d)$, and for $d \in [0.75, 4]$.

Step 1: Simulate a GARCH(p, q) model 1000 times for different samples of size (n), note also that d_{opt} will be equal to the parameter d.
In the first time we consider a GARCH$(1, 1)$ model of the form

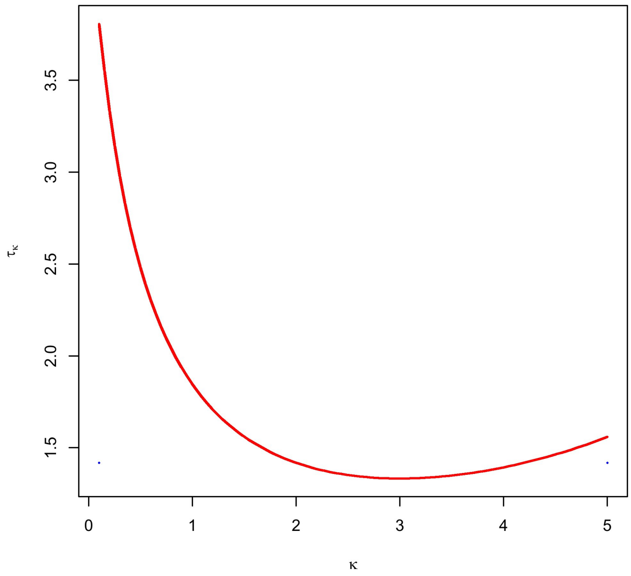


Figure 3.4 – *Optimal power κ_{opt} when the distribution of η_t is the $dgG(\frac{1}{\sqrt{2}}, 1, 3)$, where τ_ϕ is blue and τ_{GED} is red.*

$$\epsilon_t = \sigma_t \eta_t, \qquad \sigma_t^2 = 1 + 0.2\epsilon_{t-1}^2 + 0.7\sigma_{t-1}^2. \tag{3.12}$$

Then, we simulate a nested version of GARCH(p, q) model which is ARCH(1) model of the form

$$\epsilon_t = \sigma_t \eta_t, \qquad \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2. \tag{3.13}$$

We set the intercept $\omega = 1$ and to maintain the strict stationarity for any values of d, we define $\alpha = \frac{exp(-E(log\eta_1^2))}{5} = 0.95$.

Step 2: We estimate the parameters of models (3.12) and (3.13) using the Gaussian QMLE. In this step we will obtain $\hat{\eta}_t$ for each iteration $\{\hat{\eta}_1, \ldots, \hat{\eta_{1000}}\}$.

Step 3: We estimate the test statistic for $\underline{\alpha} = 5\%$ given by Equation (3.10) for each iteration which depends only on η_t.

Step 4: We test if Equation (3.10) is valid, then we obtain the frequencies of rejection of the null hypothesis.

Tables 3.1 and 3.2 gives us a clear idea concerning the frequencies of rejection of H_0. For different values of d, the gQMLE outperforms the Gaussian QMLE for the two specifications.

Table 3.1 – *Frequencies of rejection of H_0 when we consider GARCH$(1,1)$ model.*

d	0.75	1	1.4	1.7	2	2.4	2.7	3	3.5	4
n=500	0.961	0.890	0.600	0.396	0.327	0.449	0.612	0.747	0.909	0.979
n=1000	0.997	0.979	0.733	0.384	0.172	0.370	0.605	0.820	0.970	1.000
n=5000	Na	Na	0.999	0.710	0.060	0.264	0.778	0.992	1.000	1.000
n=10000	Na	Na	1	0.903	**0.039**	0.218	0.918	0.999	1	1

Table 3.2 – *Frequencies of rejection of H_0 when we consider ARCH(1) model.*

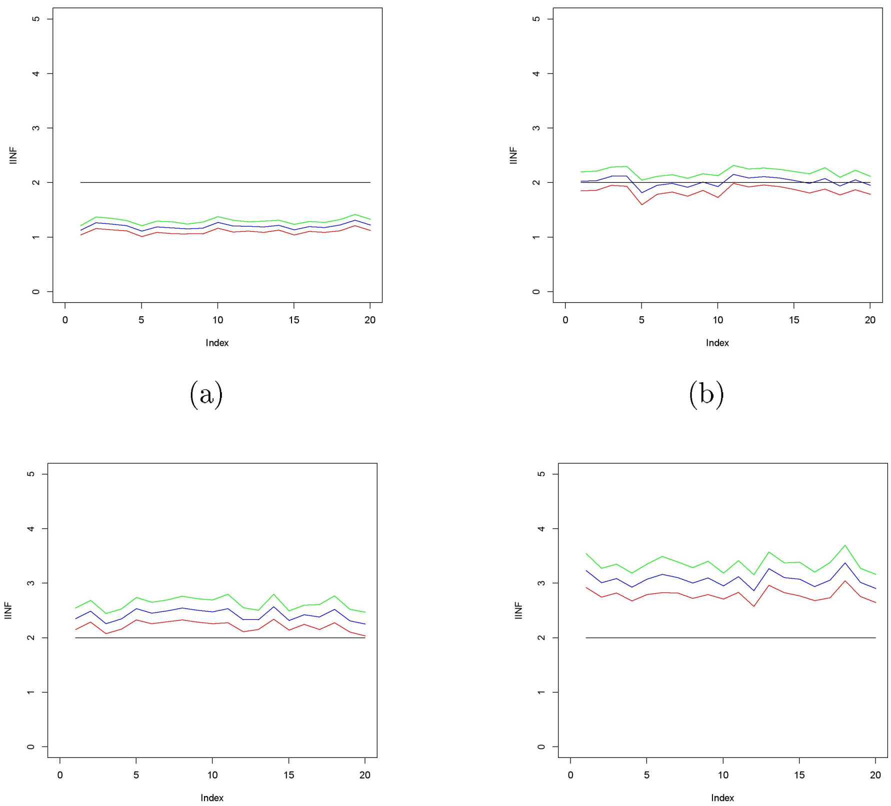
d	0.75	1	1.4	1.7	2	2.4	2.7	3	3.5	4
n=500	0.973	0.888	0.632	0.414	0.321	0.438	0.602	0.743	0.908	0.981
n=1000	0.998	0.979	0.753	0.397	0.176	0.355	0.583	0.824	0.973	1.000
n=5000	Na	1.000	0.999	0.714	0.059	0.255	0.781	0.991	1.000	1.000
n=10000	Na	1.000	1.000	0.911	**0.039**	0.209	0.916	1.000	1.000	1.000

From Tables 3.1 and 3.2, we deduce that the Gaussian QMLE loses its optimality when the density of innovations is not gaussian and this is for GARCH(p,q) models. In the case of $n = 10000$, the frequency of rejection denoted by X follows a Binomial distribution with parameters (n,p), where $n = 10000$ and $p = 0.05$. Only in this case H_0 is accepted since $P(37 \leq X \leq 64) = 0.95$.

3.5.2 Test of optimality of the Gaussian QMLE when d_{opt} is unknown

In this section we give an explicit form of the confidence interval I_n of the parameter d for GARCH$(1,1)$ and ARCH(1) models.

In the first step, we simulate a GARCH(p,q) model where the number of sample path is $N = 20$ in order to have a clear representation of the confidence interval. We consider as cited before in the simulation case when d_{opt} is known the same GARCH$(1,1)$ and ARCH(1) models with $\eta_t \sim dgG(\frac{1}{\sqrt{2}}, 1, d)$. In Step 2 we estimate the parameters of GARCH(p,q) model using the Gaussian QMLE. In Step 3, we estimate the parameter d $(\hat{d}_n)$, then we estimate the variance covariance matrix Υ_{d_n}. Finally, in Step 4, we determine the confidence interval I_n



(a) (b)

(c) (d)

Figure 3.5 – *GARCH(1,1) case: Confidence interval of* d_{opt} *when* $d = 1.2$, $d = 2$, $d = 2.4$ *and* $d = 3$ *respectively. Note that* $\hat{d}_n$ *is in blue,* $d_\phi = 2$ *is in black and lower and upper bounds are respectively in red and green.*

given by Equation (3.11).

Let's consider the case of GARCH$(1,1)$ model. First, we set $d = 1.2$ and we simulate 20 series of size 15000. Figure 3.5 (a) shows that for $d = 1.2$, $\hat{d}_n$ is very close to the true value and far from 2. This indicates that the Gaussian QMLE is not adequate for VaR parameters estimation.

In a second step, we fix $d = 2$. Figure 3.5 (b) represents the confidence interval of the estimated parameter $\hat{d}_n$ that coincides with the true value 2 in most of cases. Thus, we conclude that the Gaussian QMLE does a good job when the DGP coincides with the Gaussian distribution.

Then, we consider the case when $d = 2.4$ near to 2, according to Figure 3.5 (c) we notice that the confidence interval does not contain $d = 2$. Thereby, it is

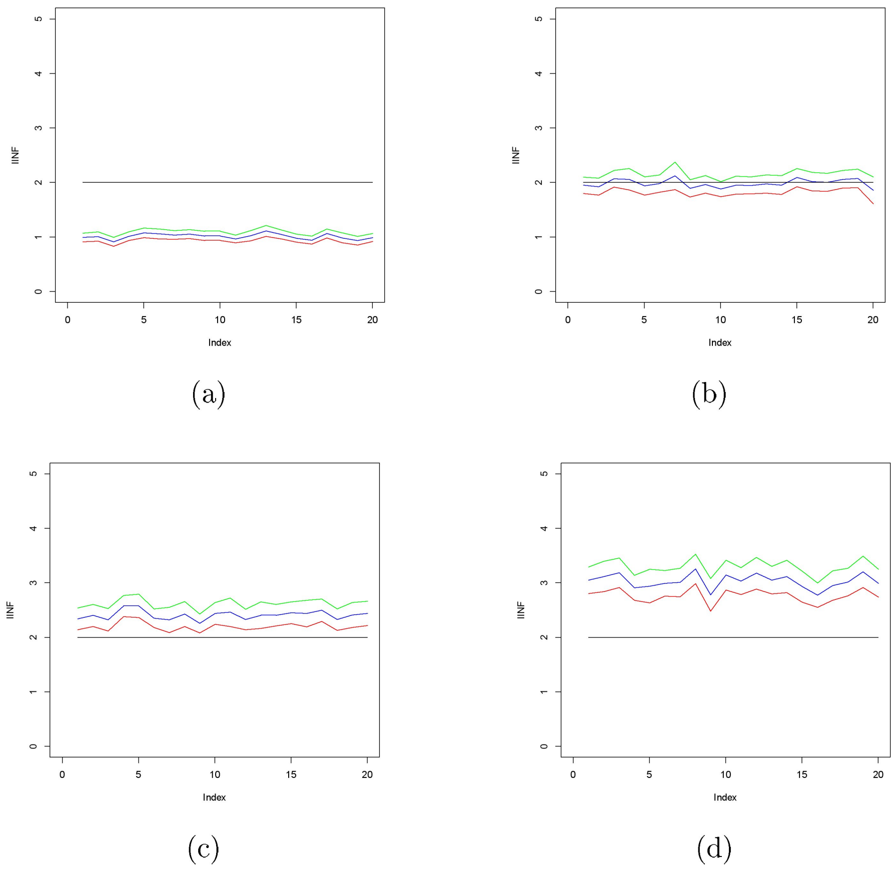


Figure 3.6 – *ARCH*(1) *case: Confidence interval of* d_{opt} *when* $d = 1$, $d = 2$, $d = 2.4$ *and* $d = 3$ *respectively. Note that* $\hat{d}_n$ *is in blue,* $d_\phi = 2$ *is in black and lower and upper bounds are respectively in red and green.*

inappropriate to use the Gaussian QMLE in such case.

Finally, considering $d = 3$, we notice that the value of $d = 2$ does not belong to the confidence interval as shown in Figure 3.5 (d). Hence, we conclude in this case that the Gaussian QMLE is not performant.

Similarly, we draw the same conclusions for ARCH(1) model with $\eta_t \sim dgG(\frac{1}{\sqrt{2}}, 1, d)$ where, $\{d = 1, 2, 2.4, 3\}$ (See Figure 3.6).

3.6 Real case study

3.6.1 Applications to financial stock price indexes

We consider 7 stock market returns which are the CAC, DAX, FTSE, Nikkei, SMI, SP500 and the TSX indices. The period of sampling spans from January 1990 to June 2013. As in the simulation study, we test the optimality of the Gaussian QMLE however here we do not make any assumption concerning the real DGP. We estimate at first GARCH$(1,1)$ model for each series, thereby we obtain the residuals η_t. At this stage, we have done a test of normality in order to have a brief description concerning the dynamic of the obtained residuals. We have used the example of TSX series where we perform the Shapiro-Wilk test of normality and the p-value of the test is about $2.2e^{-16}$ which is less than 5%. Thus, the normality is rejected. This result is verified by the normal Q-Q plot of the residuals as shown in Figure 3.7.

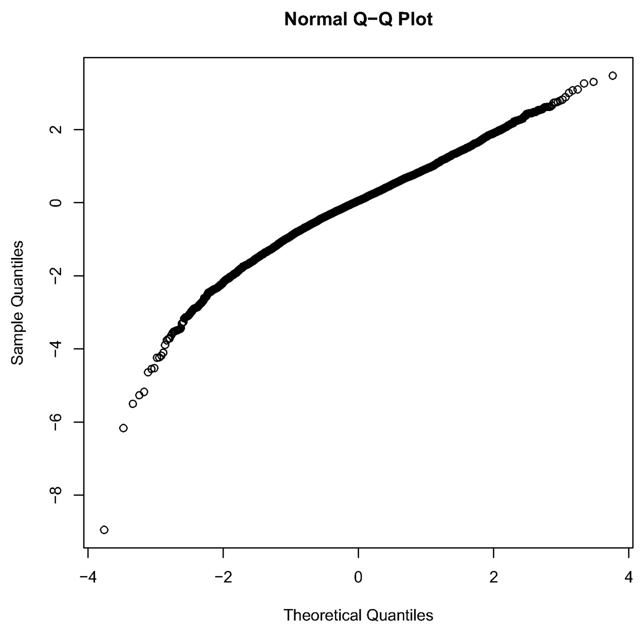


Figure 3.7 – *Normal Q-Q plot of the residuals of the TSX return series.*

In a second step we estimate $\hat{d}_n$ as a minimizer of the VaR parameter asymptotic variance and we obtain the optimal value of τ_h. Essentially, we are interested in testing the optimality of the Gaussian QMLE, that's why we calculate the p-value of the test $H_0 : d_{opt} = 2$. Note that the null is rejected if the

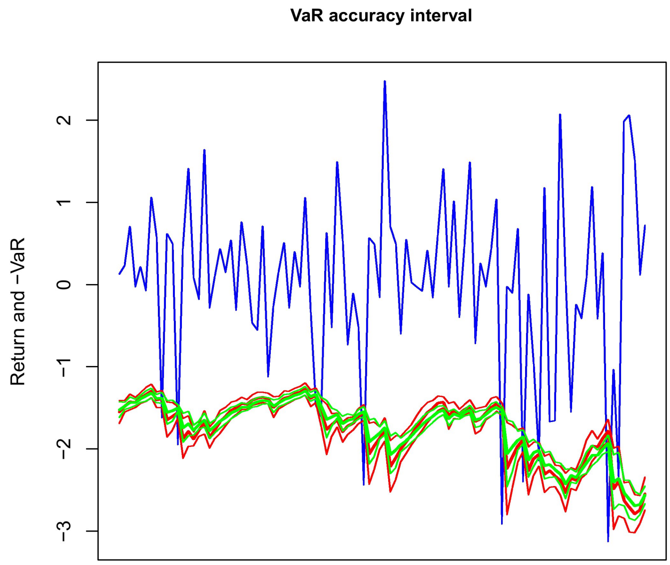
p-value is less than $\underline{\alpha}$.

Table 3.3 – *Results of the test of optimality for the Gaussian QMLE $H_0 : d = 2$ at the 5% level for 7 daily stock market returns with the estimated variance minimizer $\hat{d}_n$, the p-values, τ_h and the confidence interval I_n.*

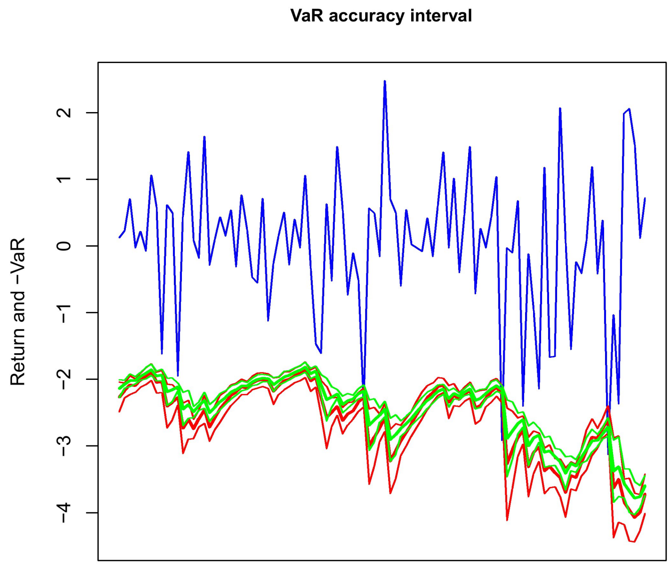
	$\hat{d}_n$	p-value	τ_h	$I_n(5\%)$
CAC40	1.172203	1.036028e-10	2.69907	$[0.9175846, 1.426821]$
DAX	0.980701	7.105427e-15	2.952382	$[0.6952428, 1.266159]$
FTSE	1.342938	1.805223e-13	2.512736	$[1.144494, 1.541381]$
NIKKEI	1.169059	0	2.803382	$[1.010572, 1.327546]$
SMI	0.9583619	1.554312e-15	2.908229	$[0.7036904, 1.213033]$
SP500	1.160155	0	2.996729	$[0.9704408, 1.349869]$
TSX	1.051079	0	2.887173	$[0.8586881, 1.24347]$

In the last step we give a confidence interval for the minimizer d_{opt} and we seek to show if the confidence interval contains the value that correspond to $d = 2$. We followed the steps cited above for two risk levels, $\underline{\alpha} = 5\%$ and $\underline{\alpha} = 1\%$. We notice from Tables 3.3 and 3.4 that all of the optimal parameter $\hat{d}_n$, τ_h and the p-value do not depend on the risk level $\underline{\alpha}$. Only the confidence interval depends on the risk $\underline{\alpha}$. Secondly, if we focus on the p-value we find that for the 7 daily stock market returns the null hypothesis is rejected. Then, the Gaussian QMLE is not appropriate for the estimation of the VaR parameter. In addition, for the 7 confidence intervals for both risk levels, we notice that no one does contain $d = 2$. Therefore, the gQMLE gains efficiency against the standard Gaussian QMLE. Hence, it is recommended to follow such test in order to verify that the choice of an instrumental density belonging to the dgG class and different from ϕ do a better job in estimating the VaR parameters.

Finally, we derive the asymptotic confidence interval of conditional VaR at 5% and 1% risk levels as shown in Figure 3.8. We used gQMLE and Gaussian QMLE for VaR parameter estimation of TSX series. Using the Equation (3.1) we find that the two confidence intervals are not confused. The confidence interval based on the gQMLE is more precise than the Gaussian QMLE. Furthermore, most of the time the confidence interval based on the gQMLE is inside that based on Gaussian QMLE.



(a)



(b)

Figure 3.8 – *The confidence interval plot of the conditional VaR for the SMI series using the Gaussian QMLE (in red) and the gQMLE based on the dgG of parameter d_{opt} (in green) for the risk levels 5% (a) and 1% (b). The estimation of the VaR parameter is based on the* 500 *previous observations.*

Table 3.4 – *Results of the test of optimality for the Gaussian QMLE $H_0 : d = 2$ at the 1% level for 7 daily stock market returns with the estimated variance minimizer $\hat{d}_n$, the p-values, τ_h and the confidence interval I_n.*

	$\hat{d}_n$	p-value	τ_h	$I_n(1\%)$
CAC40	1.172203	1.036028e-10	2.69907	$[0.8383412, 1.506064]$
DAX	0.980701	7.105427e-15	2.952382	$[0.6064012, 1.355001]$
FTSE	1.342938	1.805223e-13	2.512736	$[1.082733, 1.603142]$
NIKKEI	1.169059	0	2.803382	$[0.961247, 1.376871]$
SMI	0.9583619	1.554312e-15	2.908229	$[0.6244304, 1.292293]$
SP500	1.160155	0	2.996729	$[0.9113971, 1.408913]$
TSX	1.051079	0	2.887173	$[0.7988113, 1.303346]$

3.6.2 Applications to exchange rates and commodities

We consider daily returns of 6 exchange rates with respect to the Euro which are the USD, JPY, DKK, GBP, CHF and the CAD exchange rates. The period of sampling spans from 04 January 1999 to 17 July 2014. Then, we consider 6 daily returns of metal and oil commodities which are the Nickel, Zinc, Palladium, Gold, Silver and Oil. The data begin grom January 2010 and include the latest updates until July 2014.

Table 3.5 shows the results for the 6 exchange rates. The p-values are near zero thereby the null hypothesis is rejected. This result indicates that the Gaussian QMLE is not performant. In addition for the risk level 5% the confidence intervals does not contains $d = 2$.

Table 3.5 – *Results of the test of optimality for the Gaussian QMLE $H_0 : d = 2$ at the 5% and 1% levels for daily returns of 6 exchange rates with the estimated variance minimizer $\hat{d}_n$,τ_h and the confidence interval I_n.*

	$\hat{d}_n$	τ_h	$I_n(5\%)$	$I_n(1\%)$
USD	1.33392	2.6826	$[1.137477, 1.530363]$	$[1.076339, 1.591501]$
JPY	1.178136	2.778646	$[1.021552, 1.33472]$	$[0.9728197, 1.383453]$
DKK	0.9338318	3.683151	$[0.8358901, 1.031774]$	$[0.8054082, 1.062255]$
GBP	1.125685	2.688059	$[0.897943, 1.353427]$	$[0.8270642, 1.424306]$
CHF	0.9263633	3.451338	$[0.7959088, 1.056818]$	$[0.7553082, 1.097418]$
CAD	1.33642	2.71444	$[1.133242, 1.539598]$	$[1.070009, 1.602832]$

We draw also the confidence interval at the 1% level in order to see if it contains the value of $d = 2$ since it includes more values than the confidence interval at

5%. Following Table 3.5, all confidence intervals do not contain $d = 2$, thereby we conclude that Gaussian QMLE is also not appropriate for conditional VaR estimation as in the case of stock daily returns.

For the commodities daily returns, we reject H_0 since the p-value is less than the risk level. Table 3.6 presents the optimal value of d, τ_h and the confidence interval for both risk levels 5% and 1%.

Table 3.6 – *Results of the test of optimality for the Gaussian QMLE $H_0 : d = 2$ at the 5% and 1% levels for daily returns of 6 commodities with the estimated variance minimizer $\hat{d}_n$, τ_h and the confidence interval I_n.*

	$\hat{d}_n$	τ_h	$I_n(5\%)$	$I_n(1\%)$
Gold	0.9556055	2.985208	[0.7015983, 1.209613]	[0.622545, 1.288666]
Nickel	1.035853	3.665979	[0.753152, 1.318554]	[0.6651685, 1.406538]
Oil	1.108248	2.891139	[0.7942873, 1.422208]	[0.6965751, 1.51992]
Palladium	1.100806	2.78885	[0.8670951, 1.334517]	[0.7943585, 1.407254]
Silver	1.035853	3.665979	[0.753152, 1.318554]	[0.6651685, 1.406538]
Zinc	1.031429	3.095436	[0.8249855, 1.237873]	[0.7607352, 1.302123]

Based on the results of the three applications, we note that for both risk levels, 5% and 1%, the confidence intervals contain the value $d = 1$. This result encourages us to test if it is appropriate to recommend the use of gQMLE based on the dgG with $d = 1$ in conditional risk estimation. As given in Table 3.7, all p-values are greater than the risk levels, hence the null hypothesis stating that the dgG with $d = 1$ is useful in this task for different financial data.

Table 3.7 – *Results of the test of optimality for the gQMLE $H_0 : d = 1$*

Index	p-value	Exchange rate	p-value	Commodities	p-value
CAC40	0.5385796	USD	0.2139632	Gold	0.7122387
DAX	0.89104	JPY	0.306439	Nickel	0.8214335
FTSE	0.1162004	DKK	0.1564722	Oil	0.630048
NIKKEI	0.3299779	GBP	0.4810116	Palladium	0.5380262
SMI	0.7343694	CHF	0.2172917	Silver	0.8214335
SP500	0.4050605	CAD	0.2689116	Zinc	0.7836499
TSX	0.6410925				

3.7 Conclusion

In this section we summarize the most important steps followed to test the optimality of the Gaussian QMLE in the estimation of the VaR parameter. We

considered a large class of instrumental densities called the dgG(b, p, d) containing the Gaussian distribution for $b = 1/\sqrt{2}$, $p = 1$ and $d = 2$. In this chapter we considered GARCH(p, q) models, where we give an explicit form of the VaR parameter confidence interval based on an arbitrary instrumental density h. The Gaussian QMLE is always used, this is why we have to test the optimality of the Gaussian QMLE against the gQMLE based on the dgG distribution.
Under some assumptions we have studied the convergence and the asymptotic normality of the minimiser of the asymptotic variance $\hat{d}_n$ which depends only on the distribution of η_t. In a second step we formulated a test when d_{opt} is known. Here, testing the optimality of the Gaussian QMLE returns to estimate the minimiser of the asymptotic variance and compare it to $d_{opt} = 2$. Finally, we studied the optimality of the Gaussian QMLE when d_{opt} is unknown. In this case we provide a confidence interval for d_{opt}. We have done some illustrations showing that if we know the real DGP that we assume contains the Gaussian distribution for some defined parameter like in the case of the GED (κ), we found for the GED(2) that the minimiser $\kappa_{opt} = 2$ which coincid with the Gaussian case. However, for $\kappa \neq 2$, τ_{GED} reaches smallest values than τ_ϕ for some values near to κ_{opt}. Likewise for the case when we assumee that the real DGP is dgG $(\frac{1}{\sqrt{2}}, 1, 3)$ the Gaussian QMLE loses its optimality.
In simulation studies, for both cases, the Gaussian QMLE loses its optimality when the true DGP is dgG $(\frac{1}{\sqrt{2}}, 1, d)$ for different values of d. We have found that for 10000 simulated observations with $d = 2$ the frequencies of rejection are about 0.039 for GARCH$(1,1)$ and ARCH(1) models. These frequencies of rejection indicate that we cannot reject H_0 for a level of 5%. However, for all values of $d \neq 2$ we reject the null hypothesis, thereby we reject the optimality of the Gaussian QMLE in VaR parameters estimation for GARCH(p, q) models.

In the empirical study we estimated the variance minimizer $\hat{d}_n$, τ_h, the p-value of the test $H_0 : d_{opt} = 2$ and we proposed a confidence interval for the optimal parameter for each return series. We found that the null hypothesis is rejected for all indices. Thus, the Gaussian QMLE loses its efficiency in estimating the VaR parameters.
Finally, to reinforce the optimality of the gQMLE, we need to find a confidence interval that could contain the value of $d = 2$ for two different risk levels $\underline{\alpha} = 5\%$ and $\underline{\alpha} = 1\%$. However, it was not the case. Thereby, based on the rejection of H_0, we considered a gQMLE based on the dgG distribution of parameter d_{opt} in order to estimate the conditional VaR and establish its asymptotic confidence interval for the two risk levels.
To test whether these results can be generalized to other data, we used an application on exchange rates and commodities. We found the same results underling the

loss of efficiency of the Gaussian QMLE against the gQMLE. From these results we notice that all confidence intervals are around the value of $d = 1$. Thus, we have done a test to see if directly using the dgG with parameter $d = 1$ is more interesting than the Gaussian QMLE. Obviously, the p-values have confirmed this remark. In Future work we suggest to practitioners to apply the gQMLE based on the dgG with parameter $d = 1$.

3.8 Technical assumptions and proofs

3.8.1 Technical assumptions

The following assumptions have been used in chapter 2 in order to prove the consistency and the asymptotic normality of the conditional VaR parameter estimation using the gQMLE. For a given general volatility model of the form

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t = \sigma_t(\theta_0) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta_0) \end{cases} \tag{3.14}$$

A2: (ϵ_t) is a strictly stationary and ergodic solution of (3.14), $E|\epsilon_1|^s < \infty$ for some $s > 0$.

A3: For some $\underline{\omega} > 0$, almost surely, $\sigma_t(\theta) \in (\underline{\omega}, \infty]$ for any $\theta \in \Theta$. Moreover, for $\theta_1, \theta_2 \in \Theta$, we have $\sigma_t(\theta_1) = \sigma_t(\theta_2)$ $a.s.$ iff $\theta_1 = \theta_2$.

A4: The function $\sigma \to Eg(\eta_0, \sigma)$ is valued in $[-\infty, +\infty)$ and has a unique maximum at some point $\sigma_* \in (0, \infty)$.

A5: The instrumental density h is continuous on $\mathbb{R}$, it is also differentiable, except possibly in 0, and there exist constants $\delta \geq 0$ and $C_0 > 0$ such that, for all $u \in \mathbb{R} \setminus \{0\}$, $|uh'(u)/h(u)| \leq C_0(1 + |u|^\delta)$ and $E|\eta_0|^{2\delta} < \infty$.

A6: There exists a random variable C_1 measurable with respect to $\{\epsilon_u, u < 0\}$ and a constant $\rho \in (0, 1)$ such that $\sup_{\theta \in \Theta} |\sigma_t(\theta) - \widetilde{\sigma}_t(\theta)| \leq C_1 \rho^t$.

Under **A1** and **A4**, define the parameter

$$\theta_0^* = H(\theta_0, \sigma_*). \tag{3.15}$$

A7: The parameter θ_0^* belongs to the compact parameter space Θ.

A8: The parameter θ_0^* belongs to the interior $\overset{\circ}{\Theta}$ of Θ.

A9: There exist no non-zero $x \in \mathbb{R}^m$ such that $x' \frac{\partial \sigma_t(\theta_0^*)}{\partial \theta} = 0, \quad a.s.$

A10: The function $\theta \mapsto \sigma(x_1, x_2, \ldots; \theta)$ has continuous second-order derivatives and

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \sigma_t(\theta)}{\partial \theta} - \frac{\partial \widetilde{\sigma}_t(\theta)}{\partial \theta} \right\| + \left\| \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \widetilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq C_1 \rho^t,$$

where C_1 and ρ are as in **A6**.

A11: h is twice continuously differentiable, except possibly at 0, with $|u^2 \left(h'(u)/h(u)\right)' | \leq C_0(1 + |u|^\delta)$ for all $u \in \mathbb{R} \setminus \{0\}$ and $E|\eta_0|^\delta < \infty$, where C_0 and δ are as in **A5**.

A12: There exists a neighborhood $V(\theta_0^*)$ of θ_0^* such that

$$\sup_{\theta \in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^4, \quad \sup_{\theta \in V(\theta_0^*)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\|^2, \quad \sup_{\theta \in V(\theta_0^*)} \left| \frac{\sigma_t(\theta_0^*)}{\sigma_t(\theta)} \right|^{2\delta}$$

have finite expectations.

3.8.2 Proof of Theorem 3.1

To prove the consistency of the estimator $\hat{d}_n$, it is recommended to prove the following intermediate results under the assumptions **A1** to **A12** and **B1, B2, B3**.
Let $\mathbb{D}^* = [\underline{d}, 2\overline{d}]$, for some neighborhood $V(\theta_0^*)$ of θ_0^* we have almost surely,

$$\textbf{i)} \quad \overline{\lim_n} \sup_{s \in \mathbb{D}^*,\ \theta \in \mathcal{V}(\theta_0^*)} |\tilde{m}_n(s, 0, \theta)| < +\infty, \tag{3.16}$$

$$\textbf{ii)} \quad \overline{\lim_n} \sup_{s \in \mathbb{D}^*,\ \theta \in \mathcal{V}(\theta_0^*)} \left\| \frac{\partial \tilde{m}_n}{\partial \theta}(s, 0, \theta) \right\| < +\infty, \tag{3.17}$$

$$\textbf{iii)} \quad \overline{\lim_n} \sup_{s \in \mathbb{D}^*,\ \theta \in \mathcal{V}(\theta_0^*)} \left\| \frac{\partial \tilde{m}_n}{\partial s}(s, 0, \theta) \right\| < +\infty. \tag{3.18}$$

We need also to prove

$$\textbf{iv)} \quad \overline{\lim_n} \sup_{s \in \mathbb{D}^*,\ \theta \in \mathcal{V}(\theta_0^*)} \left\| \frac{\partial \tilde{\tau}_n}{\partial \theta}(s, \theta) \right\| < +\infty,$$

$$\textbf{v)} \quad \overline{\lim_n} \sup_{s \in \mathbb{D}^*,\ \theta \in \mathcal{V}(\theta_0^*)} \left\| \frac{\partial \tilde{\tau}_n}{\partial d}(s, \theta) \right\| < +\infty.$$

Lemma 3.2 *This Lemma is dedicated to the proof of the first three points* (**i**), (**ii**) *and* (**iii**) *by using the assumptions of Theorem (3.1).*

Proof of Lemma 3.2:

We begin to show i)*. First note that, using the ergodic theorem, which states that for a given series of iid variables $\hat{\eta}_t$ and for a given $s \in \mathbb{D}^*$ and in some neighborhood $V(\theta_0^*)$ of θ_0^*, $\overline{\lim_n} \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \frac{1}{n}\sum_{t=1}^{n}\frac{|\epsilon_t|^s}{\widetilde{\sigma}_t^s(\theta)} < +\infty$, iff:*

$$E \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^s}{\widetilde{\sigma}_t^s(\theta)} < +\infty \qquad (A.0)$$

*For this aim, using **A6** we have,*

$$\sup_{\theta\in\mathcal{V}(\theta_0^*)} |\frac{\sigma_t(\theta)}{\widetilde{\sigma}_t(\theta)} - 1| \leq C_1\rho_t,$$

more general for any $s \in \mathbb{D}^*$,

$$\sup_{\theta\in\mathcal{V}(\theta_0^*)} |\frac{\sigma_t^s(\theta)}{\widetilde{\sigma}_t^s(\theta)} - 1| \leq C_1\rho_t,$$

$$\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\frac{\frac{\sigma_t^s(\theta)}{\epsilon_t^s}}{\frac{\widetilde{\sigma}_t^s(\theta)}{\epsilon_t^s}} - 1| \leq C_1\rho_t,$$

$$\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\frac{1}{\frac{\widetilde{\sigma}_t^s(\theta)}{\epsilon_t^s}} - \frac{\epsilon_t^s}{\sigma_t^s(\theta)}| \leq C_1\rho_t \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\frac{\epsilon_t^s}{\sigma_t^s(\theta)}|,$$

$$\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\frac{\epsilon_t^s}{\widetilde{\sigma}_t^s(\theta)} - \frac{\epsilon_t^s}{\sigma_t^s(\theta)}| \leq C_1\rho_t \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\frac{\epsilon_t^s}{\sigma_t^s(\theta)}|,$$

$$\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\frac{\epsilon_t^s}{\widetilde{\sigma}_t^s(\theta)}| \leq C_1\rho_t \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\frac{\epsilon_t^s}{\sigma_t^s(\theta)}| + \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\frac{\epsilon_t^s}{\sigma_t^s(\theta)}|.$$

Thereby, we can write

$$\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^s}{\widetilde{\sigma}_t^s(\theta)} \leq C_1 \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}. \qquad (A.1)$$

To prove that $\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}$ has a finite expectation, we have to show that,

$$E \sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} < +\infty.$$

In general, for some $0 < s < k$, $k \in \mathbb{R}$ and for a given random variable $x_t, t = 1, \ldots, n$,

$$\begin{aligned}
E|x|^s &= E\left[|x_t|^s \mathbf{1}_{|x_t|^s\leq 1} + |x_t|^s \mathbf{1}_{|x_t|^s>1}\right] \\
&\leq E\left[|x_t|^k \mathbf{1}_{|x_t|^k\leq 1} + |x_t|^k \mathbf{1}_{|x_t|^k>1}\right] \\
&\leq 1 + E\left[|x_t|^k \mathbf{1}_{|x_t|^k>1}\right] \\
&\leq 1 + E\left[|x_t|^k\right] \qquad\qquad (A.2)
\end{aligned}$$

Thus by replacing x_t and k in (A.2) respectively by $\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}$ *and $2\overline{d}$ we obtain,*

$$E\left[\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right] \leq 1 + E\left[\sup_{\theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^{2\overline{d}}}{\sigma_t^{2\overline{d}}(\theta)}\right].$$

We have,

$$|\epsilon_t|^{2\overline{d}} = \sigma_t^{2\overline{d}}(\theta_0)|\eta_t|^{2\overline{d}}.$$

Then,

$$E\sup_{\theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^{2\overline{d}}}{\sigma_t^{2\overline{d}}(\theta)} = E\sup_{\theta\in\mathcal{V}(\theta_0^*)} \frac{\sigma_t^{2\overline{d}}(\theta_0)|\eta_t|^{2\overline{d}}}{\sigma_t^{2\overline{d}}(\theta)} < \infty,$$

since we have following **A5** *and* **A12***, $E|\eta_t|^{2\overline{d}} < \infty$ and $E\sup_{\theta\in\mathcal{V}(\theta_0^*)} \frac{\sigma_t^{2\overline{d}}(\theta_0)}{\sigma_t^{2\overline{d}}(\theta)} < \infty$.*

Thereby we obtain the following result

$$E\left[\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right] \leq 1 + E\left[\sup_{\theta\in\mathcal{V}(\theta_0^*)} \frac{|\epsilon_t|^{2\overline{d}}}{\sigma_t^{2\overline{d}}(\theta)}\right] < \infty. \qquad (A.3)$$

It follows that from (A.1) and (A.3)that (A.0) is verified.Thereby, i) *is proved.*

Now we give a proof of **ii)***. For any parameter θ belonging to the compact parameter space Θ, we have*

$$\begin{aligned}
\frac{\partial\tilde{m}_n}{\partial\theta}(s,0,\theta) &= \frac{1}{n}\sum_{t=1}^{n}|\epsilon_t|^s\left(\frac{-s\frac{\partial\tilde{\sigma}_t(\theta)}{\partial\theta}}{\tilde{\sigma}_t^s(\theta)\tilde{\sigma}_t(\theta)}\right), \\
&= \frac{-s}{n}\sum_{t=1}^{n}\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)}\frac{1}{\tilde{\sigma}_t(\theta)}\frac{\partial\tilde{\sigma}_t(\theta)}{\partial\theta}
\end{aligned}$$

thus,

$$\frac{\partial \tilde{m}_n}{\partial \theta}(s,0,\theta) = \frac{-s}{n}\sum_{t=1}^{n}\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)}\widetilde{D}_t(\theta). \tag{3.19}$$

By using (A.2), for a given r large ($r \geq 1$), with $0 < s(\frac{r+1}{r}) < k$, we have

$$E\left[\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)}\frac{|\epsilon_t|^{s\frac{r+1}{r}}}{\sigma_t^{s\frac{r+1}{r}}(\theta)}\right] < \infty.$$

Now, for $(\frac{r+1}{r}) \geq 1$ we have,

$$\left\|\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)}\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right\|_{\frac{r+1}{r}} = \left(\sum_{t=1}^{n}\left[\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)}\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right]^{\frac{r+1}{r}}\right)^{\frac{r}{r+1}}$$

$$\left\|\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)}\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right\|_{\frac{r+1}{r}} \leq \left(E\left[\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)}\frac{|\epsilon_t|^{s\frac{r+1}{r}}}{\sigma_t^{s\frac{r+1}{r}}(\theta)}\right]\right)^{\frac{r}{r+1}} < \infty. \tag{3.20}$$

Note also that for $\left\|\widetilde{D}_t(\theta)\right\|$, we have,

$$\left\|\widetilde{D}_t(\theta)\right\| = \left\|\frac{1}{\tilde{\sigma}_t(\theta)}\frac{\partial\widetilde{\sigma}_t(\theta)}{\partial\theta}\right\|$$

$$= \left\|\frac{1}{\tilde{\sigma}_t(\theta)}\frac{\partial\widetilde{\sigma}_t(\theta)}{\partial\theta} - \frac{1}{\tilde{\sigma}_t(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta} + \frac{1}{\tilde{\sigma}_t(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta}\right\|$$

$$\leq \frac{1}{\tilde{\sigma}_t(\theta)}\left\|\frac{\partial\widetilde{\sigma}_t(\theta)}{\partial\theta} - \frac{\partial\sigma_t(\theta)}{\partial\theta}\right\| + \frac{\sigma_t(\theta)}{\tilde{\sigma}_t(\theta)}\left\|D_t(\theta)\right\|.$$

*Following **A6** and **A10** we deduce that*

$$\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)}\left|\frac{\sigma_t(\theta)}{\tilde{\sigma}_t(\theta)}\right| \leq 1 + C_1\rho^t$$

and

$$\sup_{s\in\mathbb{D}^*,\theta\in\mathcal{V}(\theta_0^*)}\left\|\frac{\partial\widetilde{\sigma}_t(\theta)}{\partial\theta} - \frac{\partial\sigma_t(\theta)}{\partial\theta}\right\| \leq C_1\rho^t,$$

it follows then, that

$$\left\|\widetilde{D}_t(\theta)\right\| \leq C_1\rho^t + \left\|D_t(\theta)\right\|(1 + C_1\rho^t).$$

Thereby,

$$\left\|\frac{\partial \tilde{m}_n}{\partial \theta}(s,0,\theta)\right\| = \left\|\frac{-s}{n}\sum_{t=1}^{n}\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)}\widetilde{D}_t(\theta)\right\| \leq \frac{s}{n}\sum_{t=1}^{n}\left\|\widetilde{D}_t(\theta)\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\frac{\sigma_t^s(\theta)}{\tilde{\sigma}_t^s(\theta)}\right\|$$

therefore,

$$\left\|\frac{\partial \tilde{m}_n}{\partial \theta}(s,0,\theta)\right\| \leq \frac{s}{n}\sum_{t=1}^{n}(C_1\rho^t + \|D_t(\theta)\|)\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}(1+C_1\rho^t) \tag{3.21}$$

Equivalent to write

$$\left\|\frac{\partial \tilde{m}_n}{\partial \theta}(s,0,\theta)\right\| \leq \frac{s}{n}\sum_{t=1}^{n}\|D_t(\theta)\|\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}(1+C_1\rho^t) + \frac{s}{n}\sum_{t=1}^{n}C_1\rho^t(1+C_1\rho^t)\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} \tag{3.22}$$

Under the assumption **A12**, *for any real $r>0$, all positive integers k, there exist in $V(\theta_0^*)$, θ of the form $\theta = (\theta_{i_1},...,\theta_{i_{p+q+1}})$ for each element k, $i_1,...,i_k \in (1,...,p+q+1)$, we have*

$$E\sup_{\theta\in V(\theta_0^*)}\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial^k\sigma_t(\theta)}{\partial\theta_{i1}...\partial\theta_{ik}}\right\|^2 < \infty \tag{3.23}$$

Thereby, the last property can be valid for r,

$$E\sup_{\theta\in V(\theta_0^*)}\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial^k\sigma_t(\theta)}{\partial\theta_{i1}...\partial\theta_{ik}}\right\|^r < \infty \tag{3.24}$$

From (3.24), we have, $E\sup_{\theta\in V(\theta_0^*)}\|D_{t,k}\|^r < \infty$.

In order to show that

$$\overline{\lim_n}\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)}\left\|\frac{\partial \tilde{m}_n}{\partial \theta}(s,0,\theta)\right\| < \infty,$$

we have, to use the ergodic theorem and prove that $E\sup_{s\in\mathbb{D}^*,\ \theta\in V(\theta_0^*)}\|D_{t,k}(\theta)\|\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}$ is finite. Following the Hölder inequality, let, for $p,q>0$, $p=r+1$ thereby, $q=\frac{r+1}{r}$. It follows that,

$$E\sup_{s\in\mathbb{D}^*,\ \theta\in V(\theta_0^*)}\left\|D_{t,k}(\theta)\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right\| \leq E\sup_{\theta\in V(\theta_0^*)}\|D_{t,k}(\theta)\|_{r+1}E\sup_{s\in\mathbb{D}^*,\ \theta\in V(\theta_0^*)}\left\|\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right\|_{\frac{r+1}{r}}.$$

We obtain by (3.20) and (3.23) the following result

$$E \sup_{s\in\mathbb{D}^*,\ \theta\in V(\theta_0^*)} \left\| D_{t,k}(\theta) \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} \right\| < \infty,$$

and (3.17) is proven.
Now we give the prove of the third point (3.18). We have,

$$\tilde{m}_n(s,0,\theta) = \frac{1}{n}\sum_{t=1}^{n} \frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)},$$

thereby

$$\frac{\partial \tilde{m}_n}{\partial s}(s,0,\theta) = \frac{1}{n}\sum_{t=1}^{n} \frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)} \log\frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta)}. \tag{3.25}$$

It follows that $\frac{\partial \tilde{m}_n}{\partial s}(s,0,\theta) = \tilde{m}_n(s,1,\theta)$. *Then,*

$$\left\| \frac{\partial \tilde{m}_n}{\partial s}(s,0,\theta) \right\| \le \frac{1}{n}\sum_{t=1}^{n} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} \frac{\sigma_t^s(\theta)}{\tilde{\sigma}_t^s(\theta)} \log\frac{|\epsilon_t|}{\sigma_t(\theta)} \frac{\sigma_t(\theta)}{\tilde{\sigma}_t(\theta)}$$

$$\left\| \frac{\partial \tilde{m}_n}{\partial s}(s,0,\theta) \right\| \le \frac{1}{n}\sum_{t=1}^{n} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} \log\left(\frac{|\epsilon_t|}{\sigma_t(\theta)}\right)(1+C_1\rho^t) \tag{3.26}$$

By the ergodic theorem we have $\overline{\lim_n} \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \left\| \frac{\partial \tilde{m}_n}{\partial s}(s,0,\theta) \right\|$ *is finite iff*
$E \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \left\| \frac{\partial \tilde{m}_n}{\partial s}(s,0,\theta) \right\| < \infty$
equivalent to write $E \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \left\| \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} \log\frac{|\epsilon_t|}{\sigma_t(\theta)} \right\| < \infty$
Using the Hölder inequality we have

$$E \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \left\| \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} \log\frac{|\epsilon_t|}{\sigma_t(\theta)} \right\| \le E \sup_{s\in\mathbb{D}^*,\ \theta\in V(\theta_0^*)} \left\| \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} \right\|_{\frac{r+1}{r}} E \sup_{\theta\in V(\theta_0^*)} \left\| \log\frac{|\epsilon_t|}{\sigma_t(\theta)} \right\|_{r+1}.$$

Therefore by (3.20) the result follows.

Lemma 3.3 *This Lemma is dedicated to the proof of the last two points* (iv) *and* (v) *by using the assumptions of Theorem 3.1.*
Proof of Lemma 3.3:
Given

$$\widetilde{\tau}_n(d,\theta) = \frac{4}{d^2}\left(\frac{\widetilde{m}_n(2d,0,\theta)}{\widetilde{m}_n(d,0,\theta)^2} - 1 \right),$$

we have,

$$\frac{\partial \widetilde{\tau}_n}{\partial \theta}(s,\theta) = \frac{4}{s^2}\frac{1}{\widetilde{m}_n(s,0,\theta)^3}\left(\widetilde{m}_n(s,0,\theta)\frac{\partial \widetilde{m}_n(2s,0,\theta)}{\partial \theta} - 2\widetilde{m}_n(2s,0,\theta)\frac{\partial \widetilde{m}_n(s,0,\theta)}{\partial \theta}\right),$$

and

$$\begin{aligned}\frac{\partial \widetilde{\tau}_n}{\partial d}(s,\theta) &= \frac{-8}{s^3}\left(\frac{\widetilde{m}_n(2s,0,\theta)}{\widetilde{m}_n(s,0,\theta)^2} - 1\right) \\ &+ \frac{8}{s^2}\frac{1}{\widetilde{m}_n(s,0,\theta)^3}\left(\widetilde{m}_n(s,0,\theta)\frac{\partial \widetilde{m}_n(2s,0,\theta)}{\partial s} - \widetilde{m}_n(2s,0,\theta)\frac{\partial \widetilde{m}_n(s,0,\theta)}{\partial s}\right).\end{aligned}$$

It follows that

$$\frac{\partial \widetilde{\tau}_n}{\partial d}(s,\theta) = \frac{-8}{s^3}\left(\frac{\widetilde{m}_n(2s,0,\theta)}{\widetilde{m}_n(s,0,\theta)^2}-1\right)+\frac{8}{s^2}\frac{\widetilde{m}_n(s,0,\theta)\widetilde{m}_n(2s,1,\theta) - \widetilde{m}_n(2s,0,\theta)\widetilde{m}_n(s,1,\theta)}{\widetilde{m}_n(s,0,\theta)^3}. \tag{3.27}$$

To show that (iv) *and* (v) *are valid, we have to prove three points:*

a) $\overline{\lim_n} \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \left\|\widetilde{m}_n(s,0,\theta)\frac{\partial \widetilde{m}_n(2s,0,\theta)}{\partial \theta} - 2\widetilde{m}_n(2s,0,\theta)\frac{\partial \widetilde{m}_n(s,0,\theta)}{\partial \theta}\right\| < \infty$

b) $\overline{\lim_n} \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \|\widetilde{m}_n(s,0,\theta)\widetilde{m}_n(2s,1,\theta) - \widetilde{m}_n(2s,0,\theta)\widetilde{m}_n(s,1,\theta)\| < \infty$

c) $\inf_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \widetilde{m}_n(s,0,\theta) > C_1,$ *where* $C_1 > 0$.

By using Lemma 3.2 a and b are verified. Now for proving c, we have the following intermediate results.

The first order Taylor polynomial for $\widetilde{m}_n(s,0,\theta)$ *given* θ^* *between* θ *and* θ_0 *is*

$$\widetilde{m}_n(s,0,\theta) = \frac{1}{n}\sum_{t=1}^{n}\frac{|\epsilon_t|^s}{\widetilde{\sigma}_t^s(\theta_0)} + (\theta-\theta_0)'\frac{\partial \widetilde{m}_n(s,0,\theta^*)}{\partial \theta},$$

$$\widetilde{m}_n(s,0,\theta) = \widetilde{m}_n(s,0,\theta_0) + (\theta-\theta_0)'\frac{\partial \widetilde{m}_n(s,0,\theta^*)}{\partial \theta}. \qquad (A.4)$$

We have for $m(s,0) = E[\frac{|\epsilon_t|^s}{\sigma_t^s(\theta_0)}] > 0$,

$$\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\widetilde{m}_n(s,0,\theta_0) - m_n(s,0,\theta_0)| \le \frac{1}{n}\sum_{t=1}^{n}\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)}\left|\frac{|\epsilon_t|^s}{\widetilde{\sigma}_t^s(\theta_0)} - \frac{|\epsilon_t|^s}{\sigma_t^s(\theta_0)}\right|,$$

$$\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\tilde{m}_n(s,0,\theta_0) - m_n(s,0,\theta_0)| \leq \frac{1}{n}\sum_{t=1}^{n} \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\epsilon_t|^s |\frac{\sigma_t^s(\theta_0) - \widetilde{\sigma}_t^s(\theta_0)}{\sigma_t^s(\theta_0)\widetilde{\sigma}_t^s(\theta_0)}|.$$

We have, $\sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\sigma_t^s(\theta_0) - \widetilde{\sigma}_t^s(\theta_0)| \leq C_1\rho^t$, *thus*

$$\frac{1}{n}\sum_{t=1}^{n} |\epsilon_t|^s |\frac{\sigma_t^s(\theta_0) - \widetilde{\sigma}_t^s(\theta_0)}{\sigma_t^s(\theta_0)\widetilde{\sigma}_t^s(\theta_0)}| \leq C_1 \frac{1}{n}\sum_{t=1}^{n} \rho^t |\epsilon_t|^s.$$

Note that, $E[C_1 \sum_{t=1}^n \rho^t |\epsilon_t|^s] < \infty$, *thereby*

$$\overline{\lim_n} \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} |\tilde{m}_n(s,0,\theta_0) - m_n(s,0,\theta_0)| = 0, \quad a.s \qquad (A.5)$$

By the ergodic and stationarity properties we can write $\frac{|\epsilon_t|^s}{\sigma_t^s(\theta_0)} = \psi(\epsilon_t, \epsilon_{t-1}, ..., \epsilon_{t-n})$, *thus for* n *large enough we have*

$$m_n(s,0,\theta_0) = \frac{1}{n}\sum_{t=1}^{n} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta_0)} \to E[\frac{|\epsilon_t|^s}{\sigma_t^s(\theta_0)}], \quad a.s \qquad (A.6)$$

By using (A.5) we have

$$\frac{1}{n}\sum_{t=1}^{n} \frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta_0)} = \frac{1}{n}\sum_{t=1}^{n} \frac{|\epsilon_t|^s}{\sigma_t^s(\theta_0)} + o(1),$$

finally, we obtain that $\tilde{m}_n(s,0,\theta_0)$ *converges to* $m(s,0)$ *a.s.*

Here we note that following Equation (3.17), (A.4) and for θ *in the neighborhood of* θ_0, $\|(\theta-\theta_0)'\| \geq \epsilon$, *where* $\epsilon > 0$ *is small enough,*

$$\inf_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \widetilde{m}_n(s,0,\theta) \geqslant \inf_{s\in\mathbb{D}^*,} \widetilde{m}_n(s,0,\theta_0) - \sup_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \left\| \frac{\partial \widetilde{m}_n(s,0,\theta^*)}{\partial\theta} \right\| \|(\theta-\theta_0)'\|.$$

By the convergence almost surely (a.s) of $\widetilde{m}_n(s,0,\theta_0)$ *we have* $\widetilde{m}_n(s,0,\theta_0) > C_1$ *where* $C_1 > 0$. *Therefore we can say that*

$$\inf_{s\in\mathbb{D}^*,\ \theta\in\mathcal{V}(\theta_0^*)} \widetilde{m}_n(s,0,\theta) > C_1 \qquad (3.28)$$

Proof of the consistency

Let $d \neq d_0$ and let $V_k(d) = (d - \frac{1}{k}, d + \frac{1}{k})$ for any $k > 0$. For $s \in V_k(d)$, $\tilde{d} \in [s, d]$ and $\theta \in [\hat{\theta}_{n,\phi}, \theta_0]$. We have

$$\hat{\tau}_n(s) = \widetilde{\tau}_n(s, \hat{\theta}_{n,\phi}).$$

Using a Taylor expansion we obtain

$$\hat{\tau}_n(s) = \tilde{\tau}_n(d, \theta_0) + \frac{\partial \widetilde{\tau}_n}{\partial d}(\tilde{d}, \theta)(s - d) + \frac{\partial \widetilde{\tau}_n}{\partial \theta}(\tilde{d}, \theta)(\hat{\theta}_{n,\phi} - \theta_0). \tag{3.29}$$

Thereby, it follows that

$$\underline{lim}_n \inf_{s \in V_k(d)} \hat{\tau}_n(s) \geq \underline{lim}_n \hat{\tau}_n(d, \theta_0) - \overline{lim}_n \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \left\| \frac{\partial \widetilde{\tau}_n}{\partial \theta}(\tilde{d}, \theta) \right\| \left\| \hat{\theta}_{n,\phi} - \theta_0 \right\|$$

$$- \overline{lim}_n \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \left\| \frac{\partial \widetilde{\tau}_n}{\partial d}(\tilde{d}, \theta) \right\| |s - d|.$$

we have to show the following three points:

a) $\tilde{\tau}_n(d, \theta_0) = \tau_n(d, \theta_0) + o(1), \qquad a.s$

b) $\tau_n(d, \theta_0) \to \tau(d), \qquad a.s,$

c) $\tilde{\tau}_n(d, \theta_0) = \tau(d) + o(1), \qquad a.s.$

Let us begin with the proof of point a

$$\sup_{s \in V_k(d),\ \theta \in V(\theta_0)} |\widetilde{\tau}_n(d, \theta_0) - \tau_n(d, \theta_0)| = \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \left| \frac{4}{d^2} \left(\frac{\widetilde{m}_n(2d, 0, \theta_0)}{\widetilde{m}_n(d, 0, \theta_0)^2} - 1 \right) - \frac{4}{d^2} \left(\frac{m_n(2d, 0, \theta_0)}{m_n(d, 0, \theta_0)^2} - 1 \right) \right|,$$

$$= \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \frac{4}{d^2} \left| \frac{\widetilde{m}_n(2d, 0, \theta_0)}{\widetilde{m}_n(d, 0, \theta_0)^2} - \frac{m_n(2d, 0, \theta_0)}{m_n(d, 0, \theta_0)^2} \right|.$$

By using (A.5) we have

$$= \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \frac{4}{d^2} \left| \frac{\widetilde{m}_n(2d, 0, \theta_0) m_n(d, 0, \theta_0)^2 - m_n(2d, 0, \theta_0) \widetilde{m}_n(d, 0, \theta_0)^2}{\widetilde{m}_n(d, 0, \theta_0)^2 m_n(d, 0, \theta_0)^2} \right|,$$

$$= \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \frac{4}{d^2} \left| \frac{m_n(d, 0, \theta_0)^2 [\widetilde{m}_n(2d, 0, \theta_0) - (m_n(2d, 0, \theta_0) + o(1) m_n(2d, 0, \theta_0))]}{\widetilde{m}_n(d, 0, \theta_0)^2 m_n(d, 0, \theta_0)^2} \right|,$$

$$= \sup_{s\in V_k(d),\ \theta\in V(\theta_0)} \frac{4}{d^2}\left|\frac{(\widetilde{m}_n(2d,0,\theta_0)-m_n(2d,0,\theta_0))-o(1)m_n(2d,0,\theta_0)}{\widetilde{m}_n(d,0,\theta_0)^2}\right|,$$

$$= \sup_{s\in V_k(d),\ \theta\in V(\theta_0)} \frac{4}{d^2}\left|\frac{o(1)-o(1)m_n(2d,0,\theta_0)}{\widetilde{m}_n(d,0,\theta_0)^2}\right|.$$

Thereby,

$$\overline{lim}_n \sup_{s\in V_k(d),\ \theta\in V(\theta_0)} \frac{4}{d^2}\left|\frac{o(1)-o(1)m_n(2d,0,\theta_0)}{\widetilde{m}_n(d,0,\theta_0)^2}\right| = 0 \qquad a.s.$$

Then, a is verified.
For the second point b, we use the results of Equation (A.6). We have

$$\tau_d = \frac{4}{d^2}\left(\frac{E|\eta_1|^{2d}}{(E|\eta_1|^d)^2}-1\right) = \frac{4}{d^2}\left(\frac{m(2d,0)}{m(d,0)^2}-1\right).$$

Then, for n large enough, b is verified and we have

$$\tilde{\tau}_n(d,\theta_0) = \tau(d)+o(1), \qquad a.s. \qquad (A.7)$$

By using Corrollary 3.1, Lemma 3.3, assumption (**B1**), (A.7) and for $s \in V_k(d)$ where, $|s-d| = \rho$, given $\rho > 0$ is small enough when k is large, we obtain for n large enough

$$\underline{lim}_n \inf_{s\in V_k(d)} \hat{\tau}_n(s) \geq \tau(d)-\rho, \quad a.s, \qquad (A.8)$$

and then

$$\tau(d)-\rho > \tau(d_{opt}). \qquad (A.9)$$

Following the compactness argument, the compact set $\mathbb{D}$ is covered by the union of a neighborhood $V(d_{opt})$ of d_{opt} and the set of neighborhood of $V(d)$ satisfying (A.9). There exist a finite subcover of $\mathbb{D}$ of the form $V(d_{opt}), V(d_1), ..., V(d_k)$ where for $i = 1, ..., k$, $V(d_i)$ satisfies (A.9). It follows that for $i = 1, ..., k$, $d_i \neq d_{opt}$ and thus

$$\inf_{s\in\mathbb{D}} \widetilde{\tau}_n(s) = \min_{i=1,\dots,k} \inf_{s\in V(d_i)\cap V(d_{opt})} \widetilde{\tau}_n(s).$$

For n large enough, $\hat{d}_n$ belongs to $V(d_{opt})$ and

$$\inf_{s\in V(d_{opt})} \widetilde{\tau}_n(s) \leq \widetilde{\tau}_n(d_{opt}) \leq \tau_n(d_{opt}).$$

We have for n large enough $\widetilde{\tau}_n(d_{opt}) \to \tau_n(d_{opt}), a.s.$ and thus for $\epsilon > 0$ small enough, we obtain the consistency,

$$\overline{lim}_n|\hat{d}_n - d_{opt}| \leq \epsilon. \qquad a.s \tag{3.30}$$

Proof of the asymptotic normality

By using Equation(3.29) we obtain for $\tilde{d} \in [d_n, d_{opt}]$ and $\overline{\theta} \in [\hat{\theta}_{n,\phi}, \theta_0]$,

$$\frac{\partial\hat{\tau}_n}{\partial d}(d_n) = \frac{\partial\widetilde{\tau}_n}{\partial d}(d_{opt}, \theta_0) + \frac{\partial^2\widetilde{\tau}_n}{\partial d^2}(\tilde{d}, \overline{\theta})(d_n - d_{opt}) + \frac{\partial^2\widetilde{\tau}_n}{\partial d\partial\theta'}(\tilde{d}, \overline{\theta})(\hat{\theta}_{n,\phi} - \theta_0) = 0. \tag{3.31}$$

Thereby,

$$\begin{aligned}
\overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\|\frac{\partial\hat{\tau}_n}{\partial d}(d_n)\right\| &\leq \overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\|\frac{\partial\widetilde{\tau}_n}{\partial d}(d_{opt}, \theta_0)\right\| \\
&+ \overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\|\frac{\partial^2\widetilde{\tau}_n}{\partial d^2}(\tilde{d}, \overline{\theta})\right\| |d_n - d_{opt}| \\
&+ \overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\|\frac{\partial^2\widetilde{\tau}_n}{\partial d\partial\theta'}(\tilde{d}, \overline{\theta})\right\| \left\|\hat{\theta}_{n,\phi} - \theta_0\right\|,
\end{aligned}$$

$$\begin{aligned}
\overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\|\frac{\partial\hat{\tau}_n}{\partial d}(d_n) - \frac{\partial\widetilde{\tau}_n}{\partial d}(d_{opt}, \theta_0)\right\| &\leq \overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\|\frac{\partial^2\widetilde{\tau}_n}{\partial d^2}(\tilde{d}, \overline{\theta})\right\| |d_n - d_{opt}| \\
&+ \overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\|\frac{\partial^2\widetilde{\tau}_n}{\partial d\partial\theta}(\tilde{d}, \overline{\theta})\right\| \left\|\hat{\theta}_{n,\phi} - \theta_0\right\|.
\end{aligned}$$

We note that using Equation (3.27), $\frac{\partial^2\widetilde{\tau}_n}{\partial d^2}(\tilde{d}, \overline{\theta})$ and $\frac{\partial^2\widetilde{\tau}_n}{\partial d\partial\theta}(\tilde{d}, \overline{\theta})$ can be written as follows:

$$\begin{aligned}
\frac{\partial^2\widetilde{\tau}_n}{\partial d^2}(\tilde{d}, \overline{\theta}) &= \frac{24}{\tilde{d}^4}\left(\frac{\tilde{m}_n(2\tilde{d}, 0, \overline{\theta})}{\tilde{m}_n^2(\tilde{d}, 0, \overline{\theta})} - 1\right) \\
&- \frac{32}{\tilde{d}^3}\frac{\tilde{m}_n(2\tilde{d}, 1, \overline{\theta})\tilde{m}_n(\tilde{d}, 0, \overline{\theta}) - \tilde{m}_n(2\tilde{d}, 0, \overline{\theta})\tilde{m}_n(\tilde{d}, 1, \overline{\theta})}{\tilde{m}_n^3(\tilde{d}, 0, \overline{\theta})} \\
&+ \frac{8}{\tilde{d}^2}\frac{2\tilde{m}_n(2\tilde{d}, 2, \overline{\theta})\tilde{m}_n(\tilde{d}, 0, \overline{\theta})}{\tilde{m}_n^3(\tilde{d}, 0, \overline{\theta})} \\
&- \frac{8}{\tilde{d}^2}\frac{\tilde{m}_n(2\tilde{d}, 0, \overline{\theta})\tilde{m}_n(\tilde{d}, 2, \overline{\theta}) - \tilde{m}_n(2\tilde{d}, 1, \overline{\theta})\tilde{m}_n(\tilde{d}, 1, \overline{\theta})}{\tilde{m}_n^3(\tilde{d}, 0, \overline{\theta})} \\
&+ \frac{24}{\tilde{d}^2}\frac{\tilde{m}_n(2\tilde{d}, 0, \overline{\theta})\tilde{m}_n^2(\tilde{d}, 1, \overline{\theta}) - \tilde{m}_n(2\tilde{d}, 1, \overline{\theta})\tilde{m}_n(\tilde{d}, 0, \overline{\theta})\tilde{m}_n(\tilde{d}, 1, \overline{\theta})}{\tilde{m}_n^4(\tilde{d}, 0, \overline{\theta})},
\end{aligned} \tag{A.10}$$

note that

$$\frac{\partial^2 \tilde{m}_n}{\partial d^2}(s,0,\theta) = \tilde{m}_n(s,2,\theta).$$

and

$$\begin{aligned}
\frac{\partial^2 \widetilde{\tau}_n}{\partial d \partial \theta}(\tilde{d},\overline{\theta}) \quad = \quad & \frac{-8}{\tilde{d}^3}\frac{1}{\tilde{m}_n^3(\tilde{d},0,\overline{\theta})}\left[\frac{\partial \tilde{m}_n}{\partial \theta}(2\tilde{d},0,\overline{\theta})\tilde{m}_n(\tilde{d},0,\overline{\theta}) - 2\tilde{m}_n(2\tilde{d},0,\overline{\theta})\frac{\partial \tilde{m}_n}{\partial \theta}(\tilde{d},0,\overline{\theta})\right] \\
& + \frac{8}{\tilde{d}^2}\frac{1}{\tilde{m}_n^3(\tilde{d},0,\overline{\theta})}\left[\frac{\partial \tilde{m}_n}{\partial \theta}(2\tilde{d},1,\overline{\theta})\tilde{m}_n(\tilde{d},0,\overline{\theta}) + \tilde{m}_n(2\tilde{d},1,\overline{\theta})\frac{\partial \tilde{m}_n}{\partial \theta}(\tilde{d},0,\overline{\theta})\right] \\
& - \frac{8}{\tilde{d}^2}\frac{1}{\tilde{m}_n^4(\tilde{d},0,\overline{\theta})}\left[\frac{\partial \tilde{m}_n}{\partial \theta}(\tilde{d},1,\overline{\theta})\tilde{m}_n(2\tilde{d},0,\overline{\theta}) + \tilde{m}_n(\tilde{d},1,\overline{\theta})\frac{\partial \tilde{m}_n}{\partial \theta}(2\tilde{d},0,\overline{\theta})\right] \qquad (A.11) \\
& - \frac{24}{\tilde{d}^2}\frac{1}{\tilde{m}_n^4(\tilde{d},0,\overline{\theta})}\frac{\partial \tilde{m}_n}{\partial \theta}(\tilde{d},0,\overline{\theta})\left[\tilde{m}_n(2\tilde{d},1,\overline{\theta})\tilde{m}_n(\tilde{d},0,\overline{\theta}) - \tilde{m}_n(\tilde{d},1,\overline{\theta})\tilde{m}_n(2\tilde{d},0,\overline{\theta})\right].
\end{aligned}$$

It follows that by Lemma 3.2 and Lemma 3.3 we have

$$\overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\| \frac{\partial^2 \widetilde{\tau}_n}{\partial d^2}(\tilde{d},\overline{\theta}) \right\| < \infty,$$

and

$$\overline{lim}_n \sup_{\tilde{d}\in V_k(d),\ \theta\in V(\theta_0)} \left\| \frac{\partial^2 \widetilde{\tau}_n}{\partial d \partial \theta}(\tilde{d},\overline{\theta}) \right\| < \infty.$$

Thereby, for $\tilde{d} \in V_k(d)$ where k is large enough and in a small neighborhood of θ_0 we have for small enough $\rho > 0$

$$\overline{lim}_n \sup_{s\in V_k(d),\ \theta\in V(\theta_0)} \left\| \frac{\partial \hat{\tau}_n}{\partial d}(d_n) - \frac{\partial \widetilde{\tau}_n}{\partial d}(d_{opt},\theta_0) \right\| \leq \rho.$$

We have the almost sure convergence of $\widetilde{\tau}_n(d,\theta_0)$ to $\tau_n(d,\theta_0)$ and $\hat{\tau}_n(d) = \tilde{\tau}_n(d,\theta_0)$
it follows then

$$\frac{\partial \widetilde{\tau}_n}{\partial d}(d_{opt},\theta_0) - \frac{\partial \tau_n}{\partial d}(d_{opt},\theta_0) = o_p(\frac{1}{\sqrt{n}}). \qquad (3.32)$$

Equivalent to Equation (3.27) we can write

$$\frac{\partial \tau_n}{\partial d}(d_{opt},\theta_0) = \frac{-8}{d_{opt}^3}(\frac{m_n(2d_{opt},0,\theta_0)}{m_n(d_{opt},0,\theta_0)^2} - 1) +$$

$$\frac{8}{d_{opt}^2}\left[\frac{m_n(d_{opt},0,\theta_0)m_n(2d_{opt},1,\theta_0)}{m_n(d_{opt},0,\theta_0)^3} - \frac{m_n(2d_{opt},0,\theta_0)m_n(d_{opt},1,\theta_0)}{m_n(d_{opt},0,\theta_0)^3}\right].$$

In order to prove the asymptotic normality of d_n, we have to prove the following points:

i) $\sqrt{n}\left(\frac{\partial\widetilde{\tau_n}}{\partial d}(d_{opt},\theta_0) - \frac{\partial\tau}{\partial d}(d_{opt})\right) \to N(0,\zeta_{d_{opt}})$,

ii) $\frac{\partial^2\widetilde{\tau_n}}{\partial d^2}(\widetilde{d},\overline{\theta}) \overset{a.s}{\to} \frac{\partial^2\tau_\infty}{\partial d^2}(d_{opt},\theta_0)$,

where $\tau_\infty(s,\theta)$ is obtained by replacing $m_n(s,0,\theta)$ by $m_\infty(s,0,\theta)$ in Equation (3.7), and $m_\infty(s,0,\theta_0)$ is equal to $m(s,0)$.

We have to note that using Equation (A.10), and by replacing $\tilde{m}_n(\widetilde{d},\overline{\theta})$ by $m(d_{opt},0)$ we obtain

$$\begin{aligned}
\frac{\partial^2\tau_\infty}{\partial d^2}(d_{opt},\theta_0) &= \frac{24}{d_{opt}^4}\left(\frac{m(2d_{opt},0)}{m^2(d_{opt},0)} - 1\right) \\
&-\frac{32}{d_{opt}^3}\frac{m(2d_{opt},1)m(d_{opt},0) - m(2d_{opt},0)m(d_{opt},1)}{m(d_{opt},0)^3} \\
&+\frac{8}{d_{opt}^2}\frac{2m(2d_{opt},2)m(d_{opt},0)}{m^3(d_{opt},0)} \\
&-\frac{8}{d_{opt}^2}\frac{m(2d_{opt},0)m(d_{opt},2) - m(2d_{opt},1)m(d_{opt},1)}{m^3(d_{opt},0)} \\
&+\frac{24}{d_{opt}^2}\frac{m(2d_{opt},0)m^2(d_{opt},1) - m(2d_{opt},1)m(d_{opt},0)m(d_{opt},1)}{m^4(d_{opt},0)}. \qquad (A.12)
\end{aligned}$$

We have to note that by using Equation (3.27) we have

$$\frac{\partial\tau}{\partial d}(d_{opt}) = \frac{-8}{d_{opt}^3}\left(\frac{m(2d_{opt},0)}{m(d_{opt},0)^2} - 1\right) + \frac{8}{d_{opt}^2}\frac{m(d_{opt},0)m(2d_{opt},1,) - m(2d_{opt},0)m(d_{opt},1)}{m(d_{opt},0)^3} = 0, \qquad (3.33)$$

then,

$$\frac{8}{d_{opt}^3}\left(\frac{m(2d_{opt},0)}{m(d_{opt},0)^2} - 1\right) = \frac{8}{d_{opt}^2}\frac{m(d_{opt},0)m(2d_{opt},1,) - m(2d_{opt},0)m(d_{opt},1)}{m(d_{opt},0)^3}.$$

It follows that

$$\frac{32}{d_{opt}^4}\left(\frac{m(2d_{opt},0)}{m(d_{opt},0)^2}-1\right)=\frac{32}{d_{opt}^3}\frac{m(d_{opt},0)m(2d_{opt},1,)-m(2d_{opt},0)m(d_{opt},1)}{m(d_{opt},0)^3},$$

and

$$\left[-3\frac{m(d_{opt},1)}{m(d_{opt},0)}\right]\frac{8}{d_{opt}^3}\left(\frac{m(2d_{opt},0)}{m(d_{opt},0)^2}-1\right) =$$

$$\left[-3\frac{m(d_{opt},1)}{m(d_{opt},0)}\right]\frac{8}{d_{opt}^2}\frac{m(d_{opt},0)m(2d_{opt},1,)-m(2d_{opt},0)m(d_{opt},1)}{m(d_{opt},0)^3}.$$

Now we can use these equalities in the expression of $\frac{\partial^2\tau_\infty}{\partial d^2}(d_{opt},\theta_0)$, then we obtain

$$\begin{aligned}\frac{\partial^2\tau_\infty}{\partial d^2}(d_{opt},\theta_0) &= \frac{8}{d_{opt}^2}\frac{2m(2d_{opt},2)m(d_{opt},0)-m(2d_{opt},1)m(d_{opt},1)}{m^3(d_{opt},0)}\\ &-\frac{8}{d_{opt}^2}\frac{m(2d_{opt},0)m(d_{opt},2)}{m^3(d_{opt},0)}\\ &-\frac{8}{d_{opt}^3}\left(\frac{m(2d_{opt},0)}{m^2(d_{opt},0)}-1\right)\left(\frac{1}{d_{opt}}+3\frac{m(d_{opt},1)}{m(d_{opt},0)}\right)\end{aligned} \tag{3.34}$$

iii) $\frac{\partial^2\tilde{\tau}_n}{\partial d\partial\theta}(\tilde{d},\overline{\theta})\overset{a.s}{\to}\frac{\partial^2\tau_\infty}{\partial d\partial\theta}(d_{opt},\theta_0)=0.$

Proof of i:

Let us begin under the assumptions of Theorem 3.1 with the proof of point (i). Note that $\frac{\partial\tau}{\partial d}(d_{opt})=0$, then (i) can be expressed as

$$\sqrt{n}\frac{\partial\widetilde{\tau}_n}{\partial d}(d_{opt},\theta_0)\to N(0,\zeta_{d_{opt}}).$$

Given the vector

$$X_t=\begin{bmatrix}|\eta_t|^d\\ |\eta_t|^{2d}\\ |\eta_t|^d log|\eta_t|\\ |\eta_t|^{2d}log|\eta_t|\end{bmatrix},$$

by the multidimensional central limit theorem, the average of X_t, noted $\overline{X}_n$ will be

$$\overline{X}_n = \begin{bmatrix} m_n(d_{opt}, 0, \theta_0) \\ m_n(2d_{opt}, 0, \theta_0) \\ m_n(d_{opt}, 1, \theta_0) \\ m_n(2d_{opt}, 1, \theta_0) \end{bmatrix}.$$

Note that $E(X_t)$ is of the form

$$E(X_t) = \begin{bmatrix} m(d_{opt}, 0) \\ m(2d_{opt}, 0) \\ m(d_{opt}, 1) \\ m(2d_{opt}, 1) \end{bmatrix},$$

thus, $\sqrt{n}(\overline{X}_n - E(X_t))$ converges to a multivariate normal distribution of the form

$$\sqrt{n}(\overline{X}_n - E(X_t)) = \sqrt{n} \begin{bmatrix} m_n(d_{opt}, 0, \theta_0) - m(d_{opt}, 0) \\ m_n(2d_{opt}, 0, \theta_0) - m(2d_{opt}, 0) \\ m_n(d_{opt}, 1, \theta_0) - m(d_{opt}, 1) \\ m_n(2d_{opt}, 1, \theta_0) - m(2d_{opt}, 1) \end{bmatrix} \rightarrow N(0, \Sigma_{d_{opt}}),$$

where to establish the covariance matrix, we have to give an estimator of $cov(m(u, v, \theta_0)m(u', v', \theta_0)) = m(u + u', v + v') - m(u, v)m(u', v')$. Thus we obtain,

$$\Sigma_{d_{opt}} = \begin{bmatrix} \Sigma_{1,d_{opt}} \\ \Sigma_{2,d_{opt}} \\ \Sigma_{3,d_{opt}} \\ \Sigma_{4,d_{opt}} \end{bmatrix},$$

where,

$$\Sigma'_{1,d_{opt}} = \begin{bmatrix} m(2d_{opt}, 0) - m(d_{opt}, 0)^2 \\ m(3d_{opt}, 0) - m(d_{opt}, 0)m(2d_{opt}, 0) \\ m(2d_{opt}, 1) - m(d_{opt}, 0)m(d_{opt}, 1) \\ m(3d_{opt}, 1) - m(d_{opt}, 0)m(2d_{opt}, 1) \end{bmatrix},$$

$$\Sigma'_{2,d_{opt}} = \begin{bmatrix} m(3d_{opt}, 0) - m(d_{opt}, 0)m(2d_{opt}, 0) \\ m(4d_{opt}, 0) - m(2d_{opt}, 0)^2 \\ m(3r, 1) - m(2r, 0)m(r, 1) \\ m(4d_{opt}, 1) - m(2d_{opt}, 1)m(2d_{opt}, 0) \end{bmatrix},$$

$$\Sigma'_{3,d_{opt}} = \begin{bmatrix} m(2d_{opt},1) - m(d_{opt},0)m(d_{opt},1) \\ m(3d_{opt},1) - m(2d_{opt},0)m(d_{opt},1) \\ m(2d_{opt},2) - m(d_{opt},1)^2 \\ m(3d_{opt},2) - m(d_{opt},1)m(2d_{opt},1) \end{bmatrix},$$

$$\Sigma'_{4,d_{opt}} = \begin{bmatrix} m(3d_{opt},1) - m(d_{opt},0)m(2d_{opt},1) \\ m(4d_{opt},1) - m(2d_{opt},1)m(2d_{opt},0) \\ m(3d_{opt},2) - m(d_{opt},1)m(2d_{opt},1) \\ m(4d_{opt},2) - m(2d_{opt},1)^2 \end{bmatrix}.$$

Thereby using the delta method we have

$$\frac{\partial \widetilde{\tau}_n}{\partial d}(d_{opt},\theta_0) = \frac{\partial \tau}{\partial d}(d_{opt}) + \Delta'_{d_{opt}} \begin{bmatrix} m_n(d_{opt},0,\theta_0) - m(d_{opt},0) \\ m_n(2d_{opt},0,\theta_0) - m(2d_{opt},0) \\ m_n(d_{opt},1,\theta_0) - m(d_{opt},1) \\ m_n(2d_{opt},1,\theta_0) - m(2d_{opt},1) \end{bmatrix},$$

thus,

$$Var(\frac{\partial \widetilde{\tau}_n}{\partial d}(d_{opt},\theta_0)) = \Delta'_{d_{opt}} \frac{\Sigma_{d_{opt}}}{n} \Delta_{d_{opt}}$$

and the the asymptotic variance is of the form

$$\zeta_{d_{opt}} = \Delta'_{d_{opt}} \Sigma_{d_{opt}} \Delta_{d_{opt}} \tag{3.35}$$

Now we have to give an explicit form of $\Delta_{d_{opt}}$. Let a function $T(m(u,v))$ be

$$\frac{\partial \tau}{\partial d}(d_{opt}) = T \begin{bmatrix} m(d_{opt},0) \\ m(2d_{opt},0) \\ m(d_{opt},1) \\ m(2d_{opt},1) \end{bmatrix}.$$

Thus we obtain

$$\Delta_{d_{opt}} = \begin{bmatrix} \frac{\partial T}{\partial m(d_{opt},0)} \\ \frac{\partial T}{\partial m(2d_{opt},0)} \\ \frac{\partial T}{\partial m(d_{opt},1)} \\ \frac{\partial T}{\partial m(2d_{opt},1)} \end{bmatrix} = \begin{bmatrix} \frac{16}{d_{opt}^3}\frac{m(2d_{opt},0))}{m^3(d_{opt},0)} - \frac{16}{d_{opt}^2}\frac{m(2d_{opt},1))}{m^3(d_{opt},0)} + \frac{24}{d_{opt}^2}\frac{m(d_{opt},1))m(2d_{opt},0))}{m^4(d_{opt},0)} \\ \frac{-8}{d_{opt}^3}\frac{m(d_{opt},0))}{m^3(d_{opt},0)} - \frac{8}{d_{opt}^2}\frac{m(d_{opt},1))}{m^3(d_{opt},0)} \\ \frac{-8}{d_{opt}^2}\frac{m(2d_{opt},0))}{m^3(d_{opt},0)} \\ \frac{8}{d_{opt}^2}\frac{1}{m^2(d_{opt},0)} \end{bmatrix}$$

Now it is essential to show that the covariance matrix $\Sigma_{d_{opt}}$ is positive definite returns to show if $\Sigma_{d_{opt}}$ is a singular matrix or not. Let, $\lambda = \lambda_1, \lambda_2, \lambda_3, \lambda_4 \neq 0$ and a constant λ_k, we have,

$$\lambda' cov(X_t, X_t')\lambda = var(\lambda X_t),$$

when $var(\lambda X_t) = 0$ this implies that $\lambda X_t = \lambda_k \qquad a.s.$

Thus,

$$f(\eta_1) = \lambda_1 |\eta_1|^d + \lambda_2 |\eta_1|^{2d} + \lambda_3 |\eta_1|^d log|\eta_1| + \lambda_4 |\eta_1|^{2d} log|\eta_1| - \lambda_k = 0 \quad a.s.$$

Let $z = |\eta_t|^r$, then we have $\psi(z) = f(\eta_1)$, thereby we obtain

$$\psi(z) = \lambda_1 z + \lambda_2 z^2 + \lambda_3 z \frac{1}{d} log(z) + \lambda_4 z^2 \frac{1}{d} log(z) - \lambda_k = 0 \quad a.s.$$

If we differentiate $\psi(z)$ three times we obtain

$$\frac{\partial^3 \psi(z)}{\partial z} = \frac{-\lambda_3}{dz^2} + \frac{2\lambda_4}{dz},$$

thus, $\frac{\partial^3 \psi(z)}{\partial z} = 0$ for $z = \frac{\lambda_3}{2\lambda_4}$. It follows that $\psi(z)$ is null for four values contrary to what we assume in **B2**.

Proof of ii:

We have to show that

$$\frac{\partial^2 \tilde{\tau}_n}{\partial d^2}(\tilde{d}, \overline{\theta}) - \frac{\partial^2 \tilde{\tau}_n}{\partial d^2}(d_{opt}, \theta_0) = o(1). \qquad a.s \qquad (A.13)$$

Given Equation (3.29), we obtain by a Taylor expansion for $s \in [\tilde{d}, d_{opt}]$ and θ between $\overline{\theta}$ and θ_0,

$$\frac{\partial^2 \tilde{\tau}_n}{\partial d^2}(\tilde{d}, \overline{\theta}) = \frac{\partial^2 \tilde{\tau}_n}{\partial d^2}(d_{opt}, \theta_0) + \frac{\partial^3 \tilde{\tau}_n}{\partial d^3}(s, \theta)(\tilde{d} - d_{opt}) + \frac{\partial^3 \tilde{\tau}_n}{\partial d^2 \partial \theta}(s, \theta)(\overline{\theta} - \theta_0).$$

Thereby

$$\begin{aligned}
\overline{lim}_n \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \left\| \frac{\partial^2 \tilde{\tau}_n}{\partial d^2}(\tilde{d}, \overline{\theta}) \right\| \leq\ & \overline{lim}_n \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \left\| \frac{\partial^2 \tilde{\tau}_n}{\partial d^2}(d_{opt}, \theta_0) \right\| \\
& + \overline{lim}_n \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \left\| \frac{\partial^3 \tilde{\tau}_n}{\partial d^3}(s, \theta)(\tilde{d} - d_{opt}) \right\| |\tilde{d} - d_{opt}| \\
& + \overline{lim}_n \sup_{s \in V_k(d),\ \theta \in V(\theta_0)} \left\| \frac{\partial^3 \tilde{\tau}_n}{\partial d^2 \partial \theta}(s, \theta) \right\| \left\| \overline{\theta} - \theta_0 \right\|.
\end{aligned}$$

Let the functions, $C_n^i(\tilde{d},\overline{\theta})$, $F_n^i(\tilde{d},\overline{\theta})$ for $i=1,...,n$, $n\in\mathbb{N}$ and by using (A.10) we obtain,

$$\begin{aligned}\frac{\partial^3\tilde{\tau}_n}{\partial d^2\partial\theta}(\tilde{d},\overline{\theta}) &= \frac{24}{\tilde{d}^4}\frac{C_n^1(\tilde{d},\overline{\theta})}{\tilde{m}_n^4(\tilde{d},0,\overline{\theta})}-\frac{32}{\tilde{d}^3}\frac{C_n^2(\tilde{d},\overline{\theta})}{\tilde{m}_n^6(\tilde{d},0,\overline{\theta})}+\frac{8}{\tilde{d}^2}\frac{C_n^3(\tilde{d},\overline{\theta})}{\tilde{m}_n^6(\tilde{d},0,\overline{\theta})}\\ &-\frac{8}{\tilde{d}^2}\frac{C_n^4(\tilde{d},\overline{\theta})}{\tilde{m}_n^6(\tilde{d},0,\overline{\theta})}+\frac{24}{\tilde{d}^2}\frac{C_n^5(\tilde{d},\overline{\theta})}{\tilde{m}_n^8(\tilde{d},0,\overline{\theta})}, \qquad (A.14)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^3\tilde{\tau}_n}{\partial d^3}(\tilde{d},\overline{\theta}) &= \frac{-96}{\tilde{d}^5}\frac{F_n^1(\tilde{d},\overline{\theta})}{\tilde{m}_n^2(\tilde{d},0,\overline{\theta})}+\frac{96}{\tilde{d}^5}+\frac{24}{\tilde{d}^4}\frac{F_n^2(\tilde{d},\overline{\theta})}{\tilde{m}_n^4(\tilde{d},0,\overline{\theta})}\\ &+\frac{96}{\tilde{d}^4}\frac{F_n^3(\tilde{d},\overline{\theta})}{\tilde{m}_n^3(\tilde{d},0,\overline{\theta})}-\frac{32}{\tilde{d}^3}\frac{F_n^4(\tilde{d},\overline{\theta})}{\tilde{m}_n^6(\tilde{d},0,\overline{\theta})}\\ &-\frac{16}{\tilde{d}^3}\frac{F_n^5(\tilde{d},\overline{\theta})}{\tilde{m}_n^3(\tilde{d},0,\overline{\theta})}+\frac{8}{\tilde{d}^2}\frac{F_n^6(\tilde{d},\overline{\theta})}{\tilde{m}_n^6(\tilde{d},0,\overline{\theta})}\\ &+\frac{16}{\tilde{d}^3}\frac{F_n^7(\tilde{d},\overline{\theta})}{\tilde{m}_n^3(\tilde{d},0,\overline{\theta})}-\frac{8}{\tilde{d}^2}\frac{F_n^8(\tilde{d},\overline{\theta})}{\tilde{m}_n^6(\tilde{d},0,\overline{\theta})}\\ &-\frac{48}{\tilde{d}^3}\frac{F_n^9(\tilde{d},\overline{\theta})}{\tilde{m}_n^4(\tilde{d},0,\overline{\theta})}+\frac{24}{\tilde{d}^2}\frac{F_n^{10}(\tilde{d},\overline{\theta})}{\tilde{m}_n^8(\tilde{d},0,\overline{\theta})}. \qquad (A.15)\end{aligned}$$

Note that $C_n^i(\tilde{d},\overline{\theta})$, $F_n^i(\tilde{d},\overline{\theta})$ depend on $\tilde{m}_n(u,v,\overline{\theta})$ and $\frac{\partial\tilde{m}_n(u,v,\overline{\theta})}{\partial\theta}$. Thus, by using Lemma 3.2 we have

$$\overline{\lim_n}\sup_{s\in V_k(d),\ \theta\in\mathcal{V}(\theta_0^*)}\left\|C_n^1(\tilde{d},\overline{\theta})\right\|<+\infty,$$

and

$$\overline{\lim_n}\sup_{s\in V_k(d),\ \theta\in\mathcal{V}(\theta_0^*)}\left\|F_n^1(\tilde{d},\overline{\theta})\right\|<+\infty,$$

Given Equation (3.28), (A.14),(A.15) and the same conditions of Lemma 3.3, we have

$$\overline{\lim_n}\sup_{s\in V_k(d),\ \theta\in\mathcal{V}(\theta_0^*)}\left\|\frac{\partial^3\tilde{\tau}_n}{\partial d^2\partial\theta}(s,\theta)\right\|<+\infty, \qquad (3.36)$$

and,

$$\overline{\lim_n}\sup_{s\in V_k(d),\ \theta\in\mathcal{V}(\theta_0^*)}\left\|\frac{\partial^3\tilde{\tau}_n}{\partial d^3}(s,\theta)\right\|<+\infty. \qquad (3.37)$$

Thereby, (A.13) is verified.

Now for n large enough, by the law of large numbers, $\frac{\partial^2 \tilde{\tau}_n}{\partial d^2}(d_{opt}, \theta_0)$ conerges to the expected value $\frac{\partial^2 \tau_\infty}{\partial d^2}(d_{opt}, \theta_0)$ almost surely as $n \to \infty$. Then the proof is achieved.

Proof of *iii*:
Using a Taylor expansion, we obtain for $s \in [\tilde{d}, d_{opt}]$ and θ between $\overline{\theta}$ and θ_0,

$$\frac{\partial^2 \widetilde{\tau_n}}{\partial d \partial \theta}(\tilde{d}, \overline{\theta}) = \frac{\partial^2 \tilde{\tau}_n}{\partial d \partial \theta}(d_{opt}, \theta_0) + \frac{\partial^3 \tilde{\tau}_n}{\partial d^2 \partial \theta}(s, \theta)(\tilde{d} - d_{opt}) + \frac{\partial^3 \tilde{\tau}_n}{\partial d \partial \theta \partial \theta'}(s, \theta)(\overline{\theta} - \theta_0).$$

We have

$$\begin{aligned}\overline{\lim_n} \sup_{s \in V_k(d),\ \theta \in \mathcal{V}(\theta_0^*)} \left\| \frac{\partial^2 \tilde{\tau}_n}{\partial d \partial \theta}(\tilde{d}, \overline{\theta}) - \frac{\partial^2 \tilde{\tau}_n}{\partial d \partial \theta}(d_{opt}, \theta_0) \right\| &\leq \overline{\lim_n} \sup_{s \in V_k(d),\ \theta \in \mathcal{V}(\theta_0^*)} \left\| \frac{\partial^3 \tilde{\tau}_n}{\partial d^2 \partial \theta}(s, \theta) \right\| (\tilde{d} - d_{opt}) \\ &+ \overline{\lim_n} \sup_{s \in V_k(d),\ \theta \in \mathcal{V}(\theta_0^*)} \left\| \frac{\partial^3 \tilde{\tau}_n}{\partial d \partial \theta \partial \theta'}(s, \theta) \right\| \left\| \overline{\theta} - \theta_0 \right\|.\end{aligned}$$

Let the function, $K_n^i(\tilde{d}, \overline{\theta})$ for $i = 1, ..., n$, $n \in \mathbb{N}$ and by using (A.11) we obtain

$$\begin{aligned}\frac{\partial^3 \tilde{\tau}_n}{\partial d \partial \theta \partial \theta'}(\tilde{d}, \overline{\theta}) &= \frac{-8}{\tilde{d}^3} \frac{K_n^1(\tilde{d}, \overline{\theta})}{\tilde{m}_n^6(\tilde{d}, 0, \overline{\theta})} + \frac{8}{\tilde{d}^2} \frac{K_n^2(\tilde{d}, \overline{\theta})}{\tilde{m}_n^6(\tilde{d}, 0, \overline{\theta})} - \frac{8}{\tilde{d}^2} \frac{K_n^3(\tilde{d}, \overline{\theta})}{\tilde{m}_n^8(\tilde{d}, 0, \overline{\theta})} \\ &- \frac{24}{\tilde{d}^2} \frac{K_n^4(\tilde{d}, \overline{\theta})}{\tilde{m}_n^8(\tilde{d}, 0, \overline{\theta})}. \qquad (A.16)\end{aligned}$$

Note that $k_n^i(\tilde{d}, \overline{\theta})$, depends on $\tilde{m}_n(u, v, \overline{\theta})$ and $\frac{\partial \tilde{m}_n(u,v,\overline{\theta})}{\partial \theta}$. Thus by using Lemma 3.2 we have

$$\overline{\lim_n} \sup_{s \in V_k(d),\ \theta \in \mathcal{V}(\theta_0^*)} \left\| K_n^1(\tilde{d}, \overline{\theta}) \right\| < +\infty,$$

thereby, under the same argument of Lemma 3.3 and by using Equation (3.28) we have

$$\overline{\lim_n} \sup_{s \in V_k(d),\ \theta \in \mathcal{V}(\theta_0^*)} \left\| \frac{\partial^3 \tilde{\tau}_n}{\partial d \partial \theta \partial \theta'}(\tilde{d}, \overline{\theta}) \right\| < +\infty. \qquad (3.38)$$

Using (3.36) and (3.38) we obtain the almost sure convergence of $\frac{\partial^2 \tilde{\tau}_n}{\partial d \partial \theta}(\tilde{d}, \overline{\theta})$ to $\frac{\partial^2 \tilde{\tau}_n}{\partial d \partial \theta}(d_{opt}, \theta_0)$.
By the law of large number we have

$$\frac{\partial^2 \tau_\infty}{\partial d \partial \theta}(d_{opt}, \theta_0) = \frac{\partial^2 \tilde{\tau}_n}{\partial d \partial \theta}(d_{opt}, \theta_0) + o(1), \quad a.s$$

and thus

$$\frac{\partial^2 \tau_\infty}{\partial d \partial \theta}(d_{opt}, \theta_0) = \frac{\partial^2 \tilde{\tau}_n}{\partial d \partial \theta}(\tilde{d}, \overline{\theta}) + o(1). \quad a.s$$

Let $m_{n,D}(u, v, \theta) = \frac{1}{n}\sum_{t=1}^{n} D_t(\theta)|\eta_t(\theta)|^u (log|\eta_t(\theta)|)^v$. By using Equation (3.24) we obtain the almost sure convergence of $m_{n,D}(u, v, \theta)$ to $m_D(u, v)$ where $m_D(u, v) = ED_t(\theta)m(u, v)$. By using Equation (3.19), we obtain the following equalities/

$$\begin{aligned}
&\frac{\partial m_n}{\partial \theta}(s, 0, \theta) = -s m_{n,D}(s, 0, \theta),\\
&\frac{\partial m_n}{\partial \theta}(2s, 0, \theta) = -2s m_{n,D}(s, 0, \theta),\\
&\frac{\partial m_n}{\partial \theta}(s, 1, \theta) = -s m_{n,D}(s, 1, \theta) - m_{n,D}(s, 0, \theta),\\
&\frac{\partial m_n}{\partial \theta}(2s, 1, \theta) = -2s m_{n,D}(2s, 1, \theta) - m_{n,D}(2s, 0, \theta). \qquad (A.17)
\end{aligned}$$

Thereby, by replacing the terms in (A.11) by their values in (A.17) we find the form of $\frac{\partial^2 \tau_\infty}{\partial d \partial \theta}(d_{opt}, \theta_0)$ as a function of $m_D(u, v)$. Then,

$$\frac{\partial^2 \tau_\infty}{\partial d \partial \theta}(d_{opt}, \theta_0) = \frac{\partial^2 \tilde{\tau}_n}{\partial d \partial \theta}(d_{opt}, \theta_0) + o(1), \quad a.s$$

$$\begin{aligned}
\frac{\partial^2 \tau_\infty}{\partial d \partial \theta}(d_{opt}, \theta_0) \quad = \quad & \frac{16}{d_{opt}^2}\frac{1}{m^3(d_{opt}, 0)}\left[m_D(2d_{opt}, 0)m(d_{opt}, 0) - m(2d_{opt}, 0)m_D(d_{opt}, 0)\right]\\
& + \frac{16}{d_{opt}}\frac{1}{m^3(d_{opt}, 0)}\left[m_D(d_{opt}, 0)m(2d_{opt}, 1) - m_D(2d_{opt}, 1)m(d_{opt}, 0)\right]\\
& - \frac{8}{d_{opt}}\frac{1}{m^3(d_{opt}, 0)}\left[m_D(2d_{opt}, 0)m(d_{opt}, 0) - m_D(d_{opt}, 0)m(2d_{opt}, 0)\right]\\
& + \frac{8}{d_{opt}}\frac{3m(d_{opt}, 1)}{m^3(d_{opt}, 0)}\left[m_D(2d_{opt}, 0)m(d_{opt}, 0) - m_D(d_{opt}, 0)m(2d_{opt}, 0)\right]\\
& = 0. \qquad (A.18)
\end{aligned}$$

Therefore, the proof of *iii* is achieved.

It follows then by Equation (3.31) and the results driven before the convergence in law is verified.

□

References

Berkes, I. and L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* 32, 633–655.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation. *Econometrica* 50, 987–1008.

Fan, J., Qi, L. and D. Xiu (2013) Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods.*Journal of Business and Economic Statistics* 32, 178–191.

Francq, C., Lepage, G. and J-M. Zakoïan (2011) Two-stage non Gaussian QML estimation of GARCH Models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics* 165, 246–257.

Francq, C. and J-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. and J-M. Zakoïan (2013) Optimal predictions of powers of conditionally heteroskedastic processes. *Journal of the Royal Statistical Society - Series B* 75, 345-367.

Francq, C. J-M.Zaïkoan (2012) Risk-parameter estimation in volatility models. MPRA Preprint, No.41713.

Hall, P. and Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285?-317.

Kuester, K., Mittnik, S. and M.S. Paolella (2006) Value-at-Risk predictions: A comparison of alternative strategies. *Journal of Financial Econometrics* 4, 53–89.

Lee, S.W. and B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* 10, 29–52.

Chapitre 4

On Conditional risk estimation considering model risk

Abstract. Usually, parametric procedures used for conditional variance modeling are associated with model risk. Model risk may affect the volatility and conditional value at risk estimation process either due to estimation or misspecification risks. Hence, non-parametric models can be considered as alternative models given that they do not rely on an explicit form of the volatility. In this chapter, we consider the least squares support vector regression (LS-SVR), weighted LS-SVR and Fixed size LS-SVR models in order to handle the problem of conditional risk estimation taking into account issues of model risk. A simulation study and a real application show the performance of proposed volatility and VaR models.

KEYWORDS. Conditional Value at Risk, LS-SVR, GARCH models, Model risk, Spareness.

4.1 Introduction

In the context of inferring non-linear financial time series various parametric and non-parametric models are generally used. In parametric structure general autoregressive conditional heteroscedastic (GARCH) models have been widely investigated where the functional form is assumed to be known. In non-parametric framework, machine learning techniques do not assume any functional form. Support vector machine (SVM) model is considered as a volatility model that extracts information from high dimensional market data and convenient owing to its unique non-parametric, non-assumable, noise tolerant and adaptive properties. SVM is less affected by model misspecification and provides a good generalization performance (refer to Cao and Tay (2001); Telmoudi et al. (2011) and Gavrishchaka and Banerjee (2006) among others).

The above cited models among others have been advocated to the estimation of the conditional value at risk denoted by VaR (see Dowd and Blake (2006) and Brooks and Person (2003) among others). By definition, VaR corresponds to an amount that could be lost with a specified probability if the portfolio remains unmanaged over a holding period. However, the estimation of the VaR suffers from some shortcomings such as overlooking of some stylized facts and normality assumption.

Based on the influence of asymmetric effects in the accuracy of the VaR estimates, Brooks and Person (2003) proved that models which do not allow for asymmetries in the volatility specification underestimate the true VaR. These problems come certainly from the use of inappropriate models, thus the Basel committee forces banks to assess and provide solutions for protection against model risk.

In order to avoid the problem of model risk we consider non-parametric models such as the least squares support vector regression (LS-SVR) model and its variants for volatility modeling and conditional VaR estimation. We chose the LS-SVR method referring to Hable (2012) when it is considered as an M-estimator which verifies the asymptotic normality property.

LS-SVR has a few drawbacks such as loose of spareness and robustness. For robustness a weighted version of LS-SVR (WLS-SVR) can be applied by assigning a weighting factor on the basis of error variables generated from an unweighted previous step. Moreover, the sparseness can be improved by using the fixed size-LS-SVR (FS-LS-SVR). The rest of this chapter is organized as follows. In Section 4.2, we focus on model risk in a process of volatility estimation, considered to be a serious problem stemming from the selection and estimation of the model. It may happen that the selected model does not coincide with the data generating process (DGP) leading to a misspecification risk in volatility modeling. We give also an overview of some examples of symmetric and asymmetric GARCH models. In Section 4.3, we investigate the proposed non-parametric models for volatility and VaR estimation. We give some details concerning the LS-SVR procedure. Also, we give explicit details of the proposed non-parametric volatility and conditional VaR estimation methodology. In Section 4.4, empirical investigation is conducted. Finally, Section 4.5 summarizes the results.

4.2 Parametric volatility models: Estimation and specification risk

4.2.1 Model risk

Generally, there are no rules to select a precise model from alternatives for financial risk modeling. Most approaches assume that model selection does not harm the processes of estimation and prediction. This is typically considered as an inhibition of the model risk issue. Often, this problem is not taken into consideration since financial institutions do not have a clear strategy to distinguish between candidate models used as a real DGP leading to a misspecification risk. Also, if the appropriate statistical model is used, estimation risk of the parameters can be a source of doubt. In order to take into consideration this serious type of risk, Basel Committee has required financial institutions to quantify model risk: "Banks must explicitly assess the need for valuation adjustments to reflect two forms of model risk: the model risk associated with using a possibly incorrect valuation methodology; and the risk associated with using unobservable (and possibly incorrect) calibration parameters in the valuation model".
By definition, model risk follows from the fact that the DGP is unknown. In a broader sense, this is a challenging issue regarding market dynamics evolution, where Derman (1996) stated that models give at least a rough approximation of the reality. Also, Crouhy et al. (1998) have outlined that model risk is considered as the risk induced by the specification and estimation of statistical models. In line with the definition of Crouhy et al. (1998) econometric models specification could be affected by an error resulting either from omitting some relevant explanatory variables or from an unknown functional form that relates the variables leading to an incorrect inference. For instance ARCH/GARCH type models (Engle (1982) and Bollerslev (1986)) are used for financial time series modeling. Thereby it becomes problematic either to identify the correct specification or if the true model is available as one of the candidate models. Consequently, the issue of misspecification for volatility modeling cannot be avoided since the DGP is unknown, also the distributional form of the innovations is not often Gaussian. More explicitly, asset returns do not have sufficient information answering a question which is the best volatility model to use.
It is important to note that financial time series have many stylized facts such as asymmetry, fat tailedness, leptockurtocity, non-normal distribution, non-stationarity and so on. To overcome the aforementioned problem, Nelson (1991) developed the EGARCH model and Glosten et al. (1993) introduced GJR model. These models take into consideration the leverage effect. Another way to model asymmetries is to engage in modeling thresholds like the TGARCH model

introduced by Zakoïan (1994).
Despite the existence of a large range of deterministic volatility models cited above, it is not easy to find an adequate model due to the existence of stylized facts. The inference of the above models is generally based on maximum likelihood (ML) theory. This approach is suitable for the case of normal distribution, however it is not valid in real applications. Thus, quasi maximum likelihood (QML) has been used to estimate GARCH models parameters. The principle of the QML is based on maximizing a simplified form of the log-likelihood function of a misspecified model. Model misspecification consists, in a certain way, in selecting an inappropriate functional form to capture the dependency structure between the past and future prices of financial market. In this context, QML is an adequate semi-parametric statistical technique (see Francq and Zakoïan (2012)).
Similarly, non-parametric estimators for volatility estimation such as distribution-free approach SVR technique are designed to deal with an unknown functional form based only on a training set of N input-output pairs (x_i, y_i). Hence, it overcomes the problem of misspecification and avoids serious consequences of estimation ie, if the normality assumption or other specific distributions are invalid.

Gavrichshak and Ganguli (2003) have shown that SVR works effectively with high frequency financial time series without restrictive assumptions. Gavrishchaka and Banerjee (2006) have shown that SVM based volatility model is an alternative model to the existing ones in the literature. They proved that this model is tolerant to incomplete data, thereby it may address the problem of non-stationarity of financial time series, incorporate the leverage effect and cover other general non-linear effects.

4.2.2 Standard parametric volatility models

Conditional heteroskedasticity of squared returns (y_t^2) is generally modeled using GARCH type models. Let $E_{t-1}[y_t]$ be the expected return at time t conditional on information available at time $t-1$. Then, the return series at time t can be written as

$$y_t = E_{t-1}[y_t] + \varepsilon_t. \tag{4.1}$$

It is common to assume that $E_{t-1}[y_t] = 0$ without significantly modifying the performance of the model, thereby

$$y_t = \varepsilon_t = \sigma_t \eta_t, \tag{4.2}$$

where η_t is the innovation sequence, an uncorrelated process with zero mean and unit variance. The innovations η_t could be assumed normal such that

$$f(\eta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_t^2}{2}}, \text{with} - \infty < \eta_t < +\infty,$$

or follows a student distribution with ν degrees of freedom given by

$$f(\eta) = \frac{\Gamma\frac{(\nu+1)}{2}}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1 + \frac{\eta_t^2}{\nu})^{-\frac{\nu+1}{2}}, -\infty < \eta_t < +\infty,$$

or a Generalized error distribution (GED) with a location parameter κ, a scale parameter α and shape parameter β given by

$$f(\eta) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} e^{-(\frac{|\eta_t - \kappa|}{\alpha})^\beta}, -\infty < \eta_t < +\infty$$

among other distributions.

Note also that $\sigma_t^2 > 0$ is the volatility conditional on the set of information available at time $t - 1$. With reference to the literature GARCH$(1, 1)$ model provides a simple representation of the main statistical characteristics of a return series. The structure of a volatility model can be described as follows

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{4.3}$$

where the parameters ω, α and β must satisfy $\omega > 0$, $\alpha \geq 0, \beta \geq 0$ to ensure that the conditional variance is positive and $(\alpha + \beta) \leq 1$ to ensure that the process ε_t is stationary.

The GJR model takes into account the asymmetries in the response of the conditional variance of an innovation. The principle of this model is that the dynamics of conditional variance follows a switching regime depending on the sign of past innovations. GJR$(1, 1)$ specification is given by

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2. \tag{4.4}$$

The process is well defined if $\omega > 0$, $\alpha > 0$ and $\beta > 0$, and S_{t-1}^- is a dummy variable where

$$\begin{cases} S_{t-1}^- = 1, \text{if} \quad \varepsilon_{t-1} < 0, \\ S_{t-1}^- = 0, \text{else}. \end{cases}$$

Thereby, ε_{t-1}^2 has a different impact on the conditional variance σ_t^2. So $S_{t-1}^- \varepsilon_{t-1}^2$ is the squared value of negative shocks. GJR stationarity holds if and only if

$$\alpha + \beta - 0.5\gamma < 1. \tag{4.5}$$

The EGARCH model of Nelson (1991) specifies the conditional variance in logarithmic form, where the one step ahead forecast by EGARCH$(1,1)$ is given by

$$ln\sigma_t^2 = \omega + \beta ln\sigma_{t-1}^2 + \alpha\left[\frac{|\varepsilon_{t-1}|}{\sigma t-1} - \sqrt{\frac{2}{\pi}}\right] + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}}. \tag{4.6}$$

The variance depends on both the sign and the size of ε_t. The stationary conditions are verified if $\beta < 1$. EGARCH model allows good and bad news to have different impacts on volatility because the levels of ε_t are consistent with a coefficient γ.

4.3 Non-parametric models for volatility and conditional VaR estimation

Consider a parametric regression model given by $y_i = f(\rho, x_i) + \epsilon_i$ where $i = 1, \ldots, n$, x_i are the independent variables, ρ is a vector of parameters to be estimated and the errors are assumed to be iid. Non-parametric approaches avoid the incompatibility between the model structure and the behaviour of the data. It does not assume predetermined functional form but depend on the available information. The function relating the dependent variable and independent variables $f(.)$ left unspecified where the main aim is to avoid the problem of parameters estimation then estimate $f(.)$ directly. Therefore, a non-parametric regression model can be written in a similar way to parametric model as follows $y_i = f(x_i) + \epsilon_i$.

For general non-parametric regression model of the form $Y_i = f_0(X_i) + \epsilon$, the asymptotic normality of SVMs has been investigated. Where, (X_i, Y_i) are iid with a distribution P for $i = 1, ..., n$, Y is the output variable, X is the observable input variable, f_0 is the unknown regression function and ϵ is an unobservable error term. The main objective is to estimate the function f_0. Similarly as mentioned in Hable (2012), the idea behind SVR is to estimate $f_0 : X \to Y \subset \mathbb{R}$.

Given an iid sample $D_n = (x_1, y_1), ..., (x_n, y_n)$ and certain Hilbert space H of functions, the main objective is to estimate a function $f : X \to Y \subset \mathbb{R}$ element of H. Let $L(x, y, t) : Y \times \mathbb{R} \to [0, \infty)$ be a loss function, the empirical SVM f_{L,D_n,λ_n} is the solution of a minimization of regularized problem

$$\min_{f \in H} \sum_{i=1}^{n} L(x, y, f(x_i)) + \lambda_{D_n} \|f\|_H^2 ,$$

where λ is a positive regularization term.

For the theoretical SVM, let f_{L,P,λ_0} to be the solution of a smoother approximation solving the following minimization problem for a fixed regularization parameter $\lambda_0 \in (0, \infty)$,

$$\underset{f \in H}{min} \int L(x, y, f(x_i))P(d(x, y)) + \lambda_{D_n} \|f\|_H^2 .$$

Hable (2012) outlined that the difference between empirical and theoretical SVM is asymptotically normal with rate $\sqrt{n}$ such that

$$\sqrt{n}(f_{L,D_n,\lambda_n} - f_{L,P,\lambda_0}) \to GP,$$

where GP is a tight Borel-measurable Gaussian process with zero mean $E(f, GP) = 0$ in a reproducing kernel hilbert space (RKHS).
Theoretically, for SVR model the choice of the loss function depends on the application itself. Where a loss function can be described as a good one if it generates small value of loss, meaning that $f(x)$ predicts y accurately.

SVM model based $\epsilon-$ insensitive loss function (ILF) is generally used. However, its solutions arise from complex convex quadratic optimization problem. To avoid the problem of quadratic optimization problem, LS versions of SVM have been used. The LS loss function is of the form

$$L_{LS}(y) = (f(x) - y)^2.$$

Here we give an overview concerning some investigations justifying the usefulness of LS-SVR model and its variants. There are evidences showing that the LS loss function leads to obtain estimates in an easy way and LS-SVR model is performant as standard SVM based on ILF (refer to Steinwart and Christmann (2008); Gyorfi et al.(2002) and Van Gestel et al. (2004)).

Van Gestel et al.(2001) used LS-SVR with non-linear models in order to forecast the volatility of financial time series. The performance of their proposed volatility model was compared to the standard GARCH model. They have shown that LS-SVR yields comparable results to the benchmark model.

In the case of heavy tailed innovations distribution and outliers, Suykens et al. (2002) have underlined the usefulness of the WLS-SVR model, they pointed out that robustness improves by associating weights based on the resulting innovations distribution. Hou et al. (2013) have demonstrated that the use of WLS-SVR algorithm reduces the influence of noise.

Brabanter et al. (2009) have applied FS-LS-SVR to the Wiener- Hammerstein data set. In terms of root mean squared error, the FS-LS-SVR approach provides a good test performance around $4.7 * 10^{-3}$.

4.3.1 Least squares SVR and variants

Classic Model of LS-SVR: Define X and Y random variables over the probability space (Ω, A, P) where Ω is the sample set, A is the set of events and $P : A \to \mathbb{R}$ is the probability. Let χ and Υ are respectively a closed and bounded subset of $\mathbb{R}^n$ with Borel$-\sigma-$ algebra $B(\chi)$ and closed subset of $\mathbb{R}$ with Borel$-\sigma-$ algebra $B(\Upsilon)$. Let a finite sample data set $D_n = (x_1, y_1), ..., (x_n, y_n)$ of size n denotes a realization of random variables X and Y respectively, and (X_i, Y_j) for $i = j$ drawn iid from an unknown distribution F_{XY}. The goal of studying the effect of an input X on an output variable Y is to find an optimal prediction function f for an unobserved variable y.

The LS-SVR model defines the function f as follows

$$\begin{cases} f : \chi \to \Upsilon, \\ f(x) = \omega^T \varphi(x_i) + b, \\ y_i = f(x_i) + e_i, \end{cases} \tag{4.7}$$

where $\varphi(.)$ is a non-linear transformation function which maps the data into a higher possibly infinite dimensional feature space, b is the noise process and e_i are random errors. It is important to note that the estimates given by Equation (4.7) are elements of the so-called RKHS H. Then, in the primal space, the main objective is to find the optimal parameters ω and b that minimize the empirical risk functional

$$R_{emp}(\omega, b) = \frac{1}{n} \sum_{i=1}^{n} (\omega^T \varphi(x) + b - y_i)^2, \tag{4.8}$$

under the constraint $\|\omega\|^2 \leq a$, where $a \in \mathbb{R}_+$.

Following Suykens et al. (2002), the LS-SVR optimization problem is defined as follows

$$\begin{bmatrix} \min\limits_{\omega,b,e} J_p(\omega, e), J_p(\omega, e) = \frac{1}{2}\omega^T \omega + \gamma \frac{1}{2} \sum_{t=1}^{N} e_i^2 \\ such \quad that \quad y_i = \omega^T \varphi(x_i) + b + e_i, \quad i = 1, ..., N \end{bmatrix} \tag{4.9}$$

where $\varphi(.) : \mathbb{R}^n \to \mathbb{R}^{nh}$ is a non-linear mapping in kernel space, J is a loss function, $\gamma \in \mathbb{R}^+$ is an adjustable constant considered as a penalty factor and $e_i \in \mathbb{R}$ are the error variables. To solve the optimization problem in the dual space we define the Lagrangian functional

$$L(\omega, b, e) = J_p(\omega, e) - \sum_{t=1}^{N} \alpha_i (\omega^T \varphi(x_i) + b + e_i - y_i), \tag{4.10}$$

where $\alpha_i \in \mathbb{R}$ are the Lagrange multipliers also known as support vectors. According to the KKT condition the optimality conditions are given by

$$\begin{cases} \dfrac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \displaystyle\sum_{t=1}^{N} \alpha_i \varphi(x_i), \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \displaystyle\sum_{t=1}^{N} \alpha_i = 0, \\ \dfrac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i, \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow \omega^T \varphi(x_i) + b + e_i - y_i = 0, \quad i = 1, \ldots, N \end{cases} \tag{4.11}$$

and we have,

$$\begin{bmatrix} solve \quad in \quad \alpha, b: \\ \left[\frac{0}{1_v} \middle| \frac{1_v^T}{\Omega + I/\gamma}\right] \left[\frac{b}{\alpha}\right] = \left[\frac{0}{y}\right] \end{bmatrix}, \tag{4.12}$$

where, $y = [y_1, ..., y_N]$, $I_v = [1, ..., 1]$, $\alpha = [\alpha_1, ..., \alpha_N]$ and Ω is $N \times N$ Hessian matrix. Then, the kernel trick is applied here as follows

$$\omega_{ij} = \varphi^T(x_i)\varphi(x_j) = K(x_i, x_j), \qquad i, j = 1, ..., N$$

According to Mercer's theorem, the resulting LS-SVR model for function estimation becomes

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b. \tag{4.13}$$

In practice the commonly used kernel function is the radial basis kernel function (RBF) for its better generalization, it is given by

$$K(x, x_i) = exp\left(- \|x - x_i\|^2 / 2\sigma^2\right). \tag{4.14}$$

Weighted LS-SVR: When the distribution of the error variables is not Gaussian LS-SVR may lead to less robust estimates (refer to Suykens et al. (2002)). This is due to the principle of SSE cost function which gives equal weights to errors, yielding a mix of imprecise influence for some points. WLS-SVR can be conceptually applied to find the optimal cost function for the error variables.

WLS-SVR starts from standard solutions and based on the errors distributions calculates LS-SVR weights using the previous results. In the main space $\mathbb{R}^n$ the optimization equation of WLS-SVR is described as (see Suykens et al. (2002) and Hou et al. 2013)

$$\begin{bmatrix} \underset{\omega,b,e}{min} J_p(\omega, e), J_p(\omega, e) = \frac{1}{2}\omega^T\omega + \gamma\frac{1}{2}\sum_{t=1}^{N}\nu_i e_i^2 \\ such \quad that \quad y_i = \omega^T\varphi(x_i) + b + e_i). \quad i = 1, ..., N \end{bmatrix} \tag{4.15}$$

To solve the above programming, the Lagrange function can be written as follows

$$L(\omega, b, e) = J_p(\omega, e) - \sum_{t=1}^{N}\alpha_i(\omega^T\varphi(x_i) + b + e_i - y_i)$$

According to the KKT optimization conditions, the optimality conditions in Equation (4.15) are given by (Suykens et al. (2002) and Hou et al. (2013)),

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \to \omega = \sum_{t=1}^{N}\alpha_i\varphi(x_i), \\ \frac{\partial L}{\partial b} = 0 \to \sum_{t=1}^{N}\alpha_i = 0, \\ \frac{\partial L}{\partial e_i} = 0 \to \alpha_i = \frac{\gamma\nu_i e_i}{N}, \\ \frac{\partial L}{\partial \alpha_i} = 0 \to \omega^T\varphi(x_i) + b + e_i - y_i = 0, \qquad i = 1, ..., n \end{cases} \tag{4.16}$$

After dropping the constant term ω and e_i we change Equation (4.16) to the following vector equation

$$\begin{bmatrix} solve \quad in \quad \alpha, b : \\ \left[\frac{0}{1_v}\Big|\frac{1_v^T}{\Omega + V_\gamma}\right]\left[\frac{b}{\alpha}\right] = \left[\frac{0}{y}\right] \end{bmatrix}, \tag{4.17}$$

where the diagonal matrix V_γ is given by $V_\gamma = diag\left(\left[\frac{1}{\gamma v_1}, ..., \frac{1}{\gamma v_N}\right]\right)$, $\alpha = [\alpha_1, ..., \alpha_N]^T$ and $y = [y_1, ..., y_N]^T$. By solving (4.16), the parameters α_i can be obtained. Then, the WLS-SVR model for estimation is written as:

$$f(x) = \sum_{i=1}^{N}\alpha_i K(x, x_i) + b. \tag{4.18}$$

Suykens et al. (2002) weighted the training error variables $e_i = \frac{\alpha_i}{\gamma}$ from Equation (4.13) by introducing a weight factor v_i. One common choice for v_i is given by Suykens et al. (2002) as

$$\nu_i = \begin{cases} 1 & \text{if} \quad |\frac{e_i}{\hat{s}}| < c_1 \\ \frac{c_2 - |\frac{e_i}{\hat{s}}|}{c_2 - c_1} & \text{if} \quad c_1 < |\frac{e_i}{\hat{s}}| < c_2 \\ 10^{-4} & \text{otherwise.} \end{cases} \tag{4.19}$$

Note that $\hat{s}$ is a robust estimate of the standard deviation of LS-SVR errors e_i, which denotes how much the estimated error distribution deviates from a Gaussian distribution. It can be derived by (Andrzej and Shunichi, 2006)

$$\hat{s} = 1.483 Med \left\{ |e_i - Med|e_i|| \right\}. \tag{4.20}$$

The constants c_1 and c_2 depend on the percentage of outliers in the training data set which is generally unknown, that's why Suykens et al. (2002) set $c_1 = 2.5$ and $c_2 = 3$. In Equation (4.20), we refer to the median of the observations by Med.

Fixed size LS-SVR: The FS-LS-SVR model is generally designed to enhance the sparseness property and to deal with the issues of large data sets. FS-LS-SVR solves an over-determined system of M linear equations based on the Nyström approximation which introduces sparsity in the LS-SVR model. The main issue in the primal space is to give an explicit form of the non linear mapping function φ based on the decomposition of the kernel function $K(x, x_i)$. Given a data set $\{x_i, y_i\}_{i=1}^{N}$, the eigenvalues λ_i, the eigenfunctions ϕ_i, their approximations $\hat{\lambda}_i^{(s)}$ and the eigenvectors u_i the decomposition of kernel matrix Ω with entries $K(x, x_i)$ becomes feasible. Once we have the decomposition of Ω it becomes possible to obtain the approximation $\hat{\varphi}(x)$ at any point i which can be used in the primal problem (4.7) to estimate ω and b as follows

$$\hat{\varphi}_i(x^{(\nu)}) = \frac{N}{\sqrt{\lambda_i^{(s)}}} \sum_{k=1}^{N} u_{ik} K(x_k, x^{(\nu)}). \tag{4.21}$$

The finite dimensional approximation $\hat{\varphi}(x)$ can be obtained based on quadratic Renyi entropy criteria where $M << N$ support vectors are selected leading to a sparse representation in the primal space. The selection of M support vectors amounts to maximizing the quadratic Renyi entropy $H_R = -log \int p(x)^2 dx$, with an approximation of the form $\int \hat{p}(x)^2 d(x) = \frac{1}{M^2} 1^T \Omega 1$.

The algorithm for the final implementation can be described through the following steps proposed by Espinoza et al. (2004)

Step 1: Normalize inputs and outputs to have zero mean and unit variance.

Step 2: Select an initial subsample of size M.

Step 3: Build the M-size kernel matrix and compute its eigen-decomposition.

Step 4: Build the non-linear mapping approximation for the rest of the data.

Step 5: Estimate a linear regression in primal space.

Step 6: Estimate the non-linear mapping approximation for a test data.

Step 7: Use the regression estimates with the test data non-linear mapping to produce the out of sample forecast.

4.3.2 Proposed work

In this chapter we seek to apply non-parametric procedures for volatility and conditional VaR estimation. Indeed, LS-SVR model provides good performance for volatility estimation and it verifies the property of asymptotic normality as mentioned above. However, this model suffers from some shortcomings. LS-SVR procedure is only optimal if the data are corrupted by Gaussian errors distribution. Therefore, it comes in mind that the WLS-SVR and the FS-LS-SVR models could improve the conditional VaR estimation. We have to note that up to now WLS-SVR and FS-LS-SVR models are not yet used for conditional VaR estimation. Thus, we have proposed these models as new approaches which are able to compute the VaR without assuming any hypothesis on the distribution of the data. As we have noted above, these models are distribution-free approaches. Thus, we consider these models in order to avoid the problem of model risk in VaR estimation.

Using LS-SVR the volatility can be modeled as a non-linear function F of a financial time series of returns ϵ_t as follows

$$\sigma_t^2 = F[\epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_{t-p}]. \tag{4.22}$$

Given the return series ϵ_t defined by

$$\begin{cases} \epsilon_t = 100log(\frac{P_t}{P_{t-1}}) \\ \quad \epsilon_t = \sigma_t \eta_t, \end{cases} \tag{4.23}$$

where P_t is the closing stock price at time t, (η_t) is the innovations sequence and σ_t is the standard deviation. The log linear form of (4.23) is given by

$$\begin{cases} log\epsilon_t^2 = log\sigma_t^2 + (log\eta_t^2 - \mu) + \mu \\ \quad log\epsilon_t^2 = f(\epsilon_{t-1}, ..., \epsilon_{t-p}) + z_t, \end{cases} \tag{4.24}$$

where, z_t is a sequence of iid innovations. Consequently, the estimated volatility can be given by

$$\hat{\sigma}_t^2 = e^{\hat{f}(\epsilon_{t-1},\ldots,\epsilon_{t-p})} \tag{4.25}$$

Like in the work of Gavrishchaka and Banerjee (2006), LS-SVR and variants are trained on $log(\epsilon_t^2)$ instead of ϵ_t^2 and an exponential mapping is applied to the LS-SVR output. In practice, LS-SVR model and variants requires specifications of volatility in Equation (4.22) where we suppose that $\sigma_t^2 = \epsilon_{t-1}^2$.

By using the output of Equation (4.25) and for a given risk level $\alpha \in (0,1)$, the conditional VaR of the process (ϵ_t) is as follows

$$P_{t-1}\left[\epsilon_t < -VaR_t(\alpha)\right] = \alpha, \tag{4.26}$$

where P_{t-1} denotes the historical distribution conditional on$\{\epsilon_\mu, \mu < t\}$ when (ϵ_t) satisfies Equation (4.23). The theoretical VaR is then given by

$$VaR_t(\alpha) = -\sigma_t(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-p})F_z^{-1}(\alpha), \tag{4.27}$$

where $F_z^{-1}(\alpha)$ is the probability distribution function of z_t, $F_z^{-1}(\alpha) < 0$ for small values of α.

4.4 Data and Methodology

4.4.1 Simulated Data

In order to assess model risk in the task of volatility estimation and forecasting, we compare different parametric models in terms of their fit to historical data measured by the log likelihood function (LLF) and the AIC criterion. The AIC criterion is used for the evaluation of the relative goodness of fit of statistical models.

Given k the model degrees of freedom, the AIC criterion is $AIC = -2log(LLF) + 2k$. Another procedure to assess the effectiveness of volatility models is to consider the likelihood ratio test (LRtest) based on the LLF values to compare the fit of two nested models (restricted and unrestricted models). Let GARCH(p,q) model given by

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i, \epsilon_{t_i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-1}^2,$$

where GARCH$(1,1)$ model is considered as a restricted version of GARCH(p,q) model given p and q are greater than 1. It is also a restricted version of the

GJR$(1,1)$ model for a null leverage effect. The LRtest is used as a misspecification test where the following hypotheses are stated such that

$$\begin{cases} H0: The \quad restricted \quad model \quad with \quad LLF_1 \quad is \quad the \quad best \quad fitted \quad model \\ H1: The \quad unrestricted \quad model \quad with \quad LLF_2 \quad is \quad the \quad best \quad fitted \quad model. \end{cases}$$

The test statistic is

$$\Psi_{LR} = -2ln\left[\frac{LLF_2}{LLF_1}\right] \rightarrow \chi^2(DoF),$$

where the DoF of χ^2 correspond to the number of restrictions. If Ψ_{LR} is less than the critical value $H0$ is accepted at a given confidence level indicating that the nested model is more efficient to fit the historical data.
To carry out our study, the following DGP are used in the simulation of 3 datasets of size 5000. We simulate a GARCH$(1,1)$ process with parameters, $\omega = 1$, $\alpha = 0.2$ and $\beta = 0.7$. The condition of strict stationarity, $\alpha + \beta = 0.9 < 1$ is satisfied.

We consider in the first case, a GARCH$(1,1)$ process with Gaussian errors $N(0,1)$, in the second we use student innovations distribution $St(6)$ and finally a GARCH$(1,1)$ process with GED innovations distribution $GED(6)$.

To address the issue of asymmetry and heavy tailedness we consider different GARCH models containing lagged versions of the GARCH(p,q) model.

We consider 4500 observations for the in-sample set and the rest for the out-sample set. In order to underline the problem of misspecification, we fitted the in-sample dataset using symmetric and asymmetric GARCH models in addition to lagged versions of GARCH model. The main objective here is to find if the best fitted model is the true model or not in terms of AIC and LLF values.

It is clear through Table 4.1 that for the simulated dataset using GARCH$(1,1)$ model with Gaussian innovations, the best fitted model does not coincide with the true model. Interesting results are given in Table 4.1 since the LLF and AIC values based on GED innovations for the true models are almost the best ones.

Firstly, consider only the fitted models GARCH$(1,1)$ with three different distributions, in the case of simulated data using GARCH$(1,1)$ model with Gaussian, student and GED innovations. The best fitted model is the GARCH$(1,1)$ with GED distribution (see the stared values in Table 4.1).

Then, for more general interpretation we consider only the Gaussian and student distribution where the best fitted model is almost the GARCH$(5,1)$ model.

Now consider the GED distribution, in the first case we find that the best fitted model is the GJR$(1,1)$ model based on the GED distribution. In the second and third cases, the GARCH$(3,1)$ based on GED innovations is better than the true model in terms of LLF and AIC values. It is important to note that the unrestricted versions of the GARCH(p, q) model are quite similar. However, the EGARCH$(1, 1)$ model is still different and generates the highest values of AIC. So it is clear that there are no rules for choosing the parametric models. In another terms the use of a specific version of GARCH models from many alternative models still unjustified. This underlines the problem of model risk linked to the setting of the model. In addition, the choice of the innovations' distribution denoted by h is also of great importance since h can belong to a wide large class of distributions such as the exponential family. Then, the selection of h is fixed under some hypotheses. It is adequate in this case to avoid the problem to switch to non-parametric approaches which do not assume neither any functional form of the volatility nor any hypotheses of innovations' distribution.

Table 4.1 – *LLF and AIC values of the parametric models*

Fitted		Gaussian		Student		GED	
Model		LLF	AIC	LLF	AIC	LLF	AIC
GARCH	Gaussian	-11082	22173	-10937	21881	-11342	22693
(1, 1)	Student	-11083	22176	-10811	21631	-11354	22717
	GED	**-5667***	**11345***	**-5912***	**11834***	**-5297***	**10604***
GARCH	Gaussian	-11080	22170	-10934	21879	-11340	22691
(1, 2)	Student	-11081	22173	-10808	21629	-11352	22715
	GED	-5666	11344	-5911	11834	-5296	10604
GARCH	Gaussian	**-11074**	**22163**	-10928	21871	**-11334**	**22684**
(1, 5)	Student	-11074	22166	**-10802**	**21622**	-11345	22708
	GED	-5663	11345	-5907	11833	-5293	10604
GARCH	Gaussian	-11078	22168	-10932	21877	-11338	22688
(3, 1)	Student	-11079	22171	-10806	21626	-11349	22712
	GED	-5665	11344	**-5910**	**11834**	**-5295**	**10604**
GJR	Gaussian	-11081	22171	-10937	21883	-11342	22695
(1, 1)	Student	-11081	22175	-10810	21633	-11354	22719
	GED	**-5663**	**11338**	-5912	11836	-5297	10606
EGARCH	Gaussian	-11096	22202	-10952	21913	-11348	22705
(1,1)	Student	-11096	22204	-10819	21649	-11348	22708
	GED	-5691	11394	-5936	11884	-5320	10653

Thus, the true model cannot be considered as the best fitting model. This result is confirmed by the likelihood ratio test (LR-test) shown in Table 4.2, where H denotes here the decision taken.

$$\begin{cases} H = 0 & The \quad restricted \quad model \quad is \quad the \quad best \quad fitted \quad model \\ H = 1 & Otherwise. \end{cases}$$

Table 4.2 indicates that the true model which is a restricted version of GARCH(p, q) models and GJR$(1, 1)$ model is not usually the selected model.
In the first case where the DGP coincides with a GARCH$(1, 1)$ model with Gaussian innovations, at the risk level of 5%, the true model is only efficient in two times against the GJR$(1, 1)$ model with Gaussian innovations and with student innovations. However, at the level of 1%, in most cases, the true model is less efficient than the GARCH$(1, 5)$ with different distributions and is overcome by alternative models conditional on the GED distribution.

In the student case, it is clear for the two risk levels that all the used models for different distributions, except the GJR$(1, 1)$ with Gaussian innovations and Student innovations, have better performance than the true model in fitting the data.

Finally, in the GED case, the true model is outperformed by all the models with Gaussian and student distributions. However, it has the best fitting performance when we consider only the GED distribution. Thereby, it is important to note that the model specification and the choice of the innovations distribution are critical issues in parametric volatility modelling.

4.4.2 Descriptive statistics

In this section, we give descriptive statistics of seven stock indexes, the DAX, CAC40, NIKKEI225, FTSE100, SP500, SMI and TSX. The dataset consists of daily data that encompass the period from 1991 to 2013. There are 5500 observations of daily prices which are transformed into daily returns (ϵ_t) before conducting the analysis, where it satisfies Equation (4.23).

For more details about the seven time series characteristics, their descriptive statistics are summarized in Table 4.3.

The seven series are characterized by negative skewness and kurtosis coefficients greater than the normal level. The Jarque Bera statistics give overwhelming evidence that the normality null hypothesis is rejected in all cases. This indicates that the seven time series distributions behave differently from the normal one. These results indicate that standard volatility model based on normal distribution

Table 4.2 – *LR-test at 95% and 99% confidence levels*

Gaussian		pvalue	stat	H 5 %	H 1%
GARCH	Gau	0455	4	1	0
(1, 2)	STD	0.0455	4	1	0
	GED	0	10832	1	1
GARCH	Gau	0.0030	16	1	1
(1, 5)	STD	0.0030	16	1	1
	GED	0	10838	1	1
GARCH	Gau	0.0183	8	1	0
(3, 1)	STD	0.0498	6	1	0
	GED	0	10834	1	1
GJR	Gau	0.1573	2	0	0
(1, 1)	STD	0.1573	2	0	0
	GED	0	10838	1	1
Student		pvalue	stat	H 5 %	H 1%
GARCH	Gau	0	-246	1	0
(1, 2)	STD	0.0143	6	1	1
	GED	0	9800	1	1
GARCH	Gau	0	-234	1	1
(1, 5)	STD	0.0012	18	1	1
	GED	0	9808	1	1
GARCH	Gau	0	-242	1	1
(3, 1)	STD	0.0067	10	1	1
	GED	0	9802	1	1
GJR	Gau	1	-252	0	0
(1, 1)	STD	0.1573	2	0	0
	GED	0	9798	1	1
GED		pvalue	stat	H 5 %	H 1%
GARCH	Gau	0	-12086	1	1
(1, 2)	STD	0	-12110	1	1
	GED	0.1573	2	0	0
GARCH	Gau	0	-12074	1	1
(1, 5)	STD	0	-12096	1	1
	GED	0.0916	8	0	0
GARCH	Gau	0	-12082	1	1
(3, 1)	STD	0	-12104	1	1
	GED	0.1353	4	0	0
GJR	Gau	0	-12090	1	1
(1, 1)	STD	0	-12114	1	1
	GED	1	0	0	0

Table 4.3 – *Summary descriptive statistics for the return series*

	CAC40	DAX	SP500	NIKKEI225	FTSE100	SMI	TSX
Mean	0.0056	0.0129	0.0111	-0.0050	0.0070	0.0121	0.0098
Maximum	4.6012	4.6893	4.7586	5.7477	4.0755	4.6850	4.0694
Minimum	-4.1134	-3.2283	-4.1126	-5.2598	-4.0235	-3.5212	-4.2508
Std. Dev.	0.6235	0.6339	0.5118	0.6649	0.5024	0.5109	0.4600
Skewness	-0.0137	-0.0686	-0.2442	-0.2100	-0.1131	-0.0892	-0.7208
Kurtosis	7.3871	7.3393	11.7079	8.1980	8.9847	8.6296	13.2279
Jarque Bera	44108	43194	17432	62324	82198	72700	24449

has a hard time forecasting the true volatility. Furthermore, we test for asymmetric effect on conditional volatility for the studied time series. A correlation coefficient between the squared returns and lagged returns $(corr\epsilon_t^2, \epsilon_{t-1})$ is estimated. In Table 4.4, negative coefficient values indicate the existence of a leverage effect that refers to the idea that price movement are negatively correlated with volatility, except for the SP500 time series. Table 4.4 provides evidence concerning the existence of some stylized facts such as asymmetry (negative correlation coefficient). From Table 4.4 we notice that parametric GARCH$(1,1)$ model is not appropriate. Hence, in order to take into consideration the leverage effect the use of asymmetric GARCH models could be useful in this case, like GJR, EGARCH, APARCH, Log-GARCH and TGARCH models among others. Hence, there are two inconveniences. First, there is no information to select an appropriate model from these candidate models. Second, we have the issue of the random choice of the innovations' distribution. In order to avoid the problem of model risk and misspecification risk, as cited above, we should consider non-parametric models instead of parametric ones. For this reason we give empirical evidence concerning the performance of non-parametric models in volatility modeling and conditional VaR estimation.

4.4.3 Empirical results

In this subsection we have used 5000 observations for the training set and the remaining data as the testing set. We compare the performance of three non-parametric models namely, LS-SVR, WLS-SVR and FS-LS-SVR for the out of sample volatility prediction. In our study an initial setting of the kernel parameter and gamma is done, where we fix them randomly. We set M=15 at random for the FS-LS-SVR model, knowing that there is no formal method that can be used

Table 4.4 – *Asymmetric effect test on conditional volatility*

Daily return indexes	$Corr\epsilon_t^2, \epsilon_{t-1}$
DAX	-0,07518341
CAC40	-0,06062405
FTSE	-0,03128264
SP500	0,001101030
NIKKEI225	-0,05568448
SMI	-0,05120143
TSX	-0,02675283

for its selection. By using the Corollary mentioned above, Figure 4.1 presents the forecasted volatility plot of the seven time series. It can be readily seen that in all cases, LS-SVR and WLS-SVR models behave similarly. There is a little variation between the two volatility models. However, FS-LS-SVR model provides a volatility which has the same shape as the two concurrent models but with a clear difference of the volatility values.

MSE, RMSE, and MAE performance measures: In order to check the forecasting performance of the concurrent models in the out sample set, the Mean squared error (MSE), Root mean squared error (RMSE) and Mean absolute error (MAE) are considered as goodness-of-fit measures. These criteria penalize equally over and under predictions. Generally, the MSE is more sensitive to the miss-prediction than the MAE. The best predictor is the one with the lowest values of these loss functions given by:

- $MSE = \frac{1}{N}\sum_{t=1}^{N}(\epsilon_t^2 - \hat{\sigma}_t^2)^2$

- $RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\epsilon_t^2 - \hat{\sigma}_t^2)^2}$

- $MAE = \frac{1}{N}\sum_{t=1}^{N}|\epsilon_t^2 - \hat{\sigma}_t^2|$

Table 4.5 gives a clear idea about the performance of the SVR model used in the work of Gavrishchaka and Banerjee (2006) based on the RBF kernel function and the three candidate models. By means of MSE, RMSE and MAE, we find that SVR model using the $\epsilon-$ILF loss function has the highest values of the performance criteria. FS-LS-SVR model has the best performance followed by the WLS-SVR

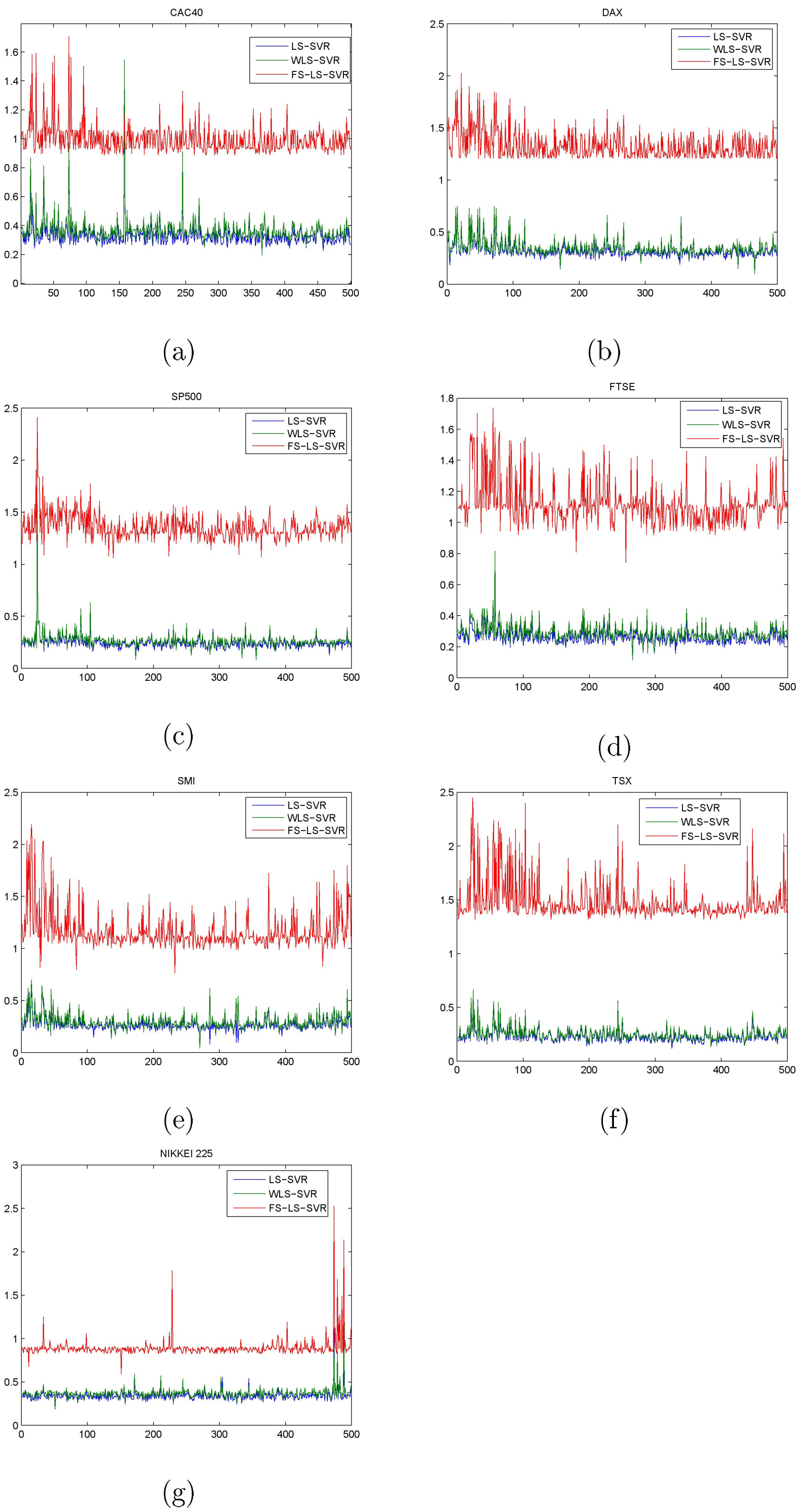


Figure 4.1 – *Volatility plot using LS-SVR, WLS-SVR and FS-LS-SVR.*

model. Here we can underline the role of the sparse approximation. Indeed, the enhancement of sparseness has a better influence then the weighting of the error variables in most cases. Here we assess the main advantages of FS-LS-SVR model. It is considered as a powerful algorithm to solve a large dataset regression task over a determined value of $M << 5000$. Thus, it is important to note that the training in a fixed subset entails a good precision. In this sense, the active selection of the training subset may be the most representative of all data.

Table 4.6 summarizes the results using LS-SVR and its variants for conditional VaR estimation given by Equation (4.26). The quantile of the residuals is given for 5% and 1% levels.

In order to compare the resulting VaR at 5% and 1% of the three models, we compare all VaR values to the mean of VaR and determine the number of violations: ie, the out-sample size (500) multiplied by the risk level. The number of violations should not exceed 25 and 5 times for 5% and 1% levels respectively. We notice that FS-LS-SVR is the adequate model with a number of violations less than the threshold for the risk level 5% and 1% . Also we note that for the TSX series, LS-SVR and WLS-SVR models are valid for the risk level 5%.

4.5 Conclusion

In this chapter we seek to comprehensively describe the estimation risk induced by standard parametric GARCH models and the ability of non-parametric models in reducing or avoiding this risk. A simulation study using three datasets based on the Gaussian, Student and GED distributions is conducted. We estimated the volatility based on six different models which contain the true DGP. Using the AIC and LLF criteria along with the LR test we have found that the best fitting model is not necessarily the true model for fitting historical data. Thereby, an alternative way to avoid the problem of model risk is to use non-parametric models for volatility and conditional VaR estimation. Using performance evaluation criteria, we have found that FS-LS-SVR model is the best model for volatility prediction followed by WLS-SVR model. Also it is well suited for conditional VaR estimation. Thus, machine learning distribution-free approaches can be considered as alternative models to parametric ones. Indeed, in financial risk management FS-LS-SVR model performance for conditional VaR estimation can be extended to other distortion risk measures such as the expected shortfall.

References

Andrzej, C. and A. Shunichi (2005) Adaptive blind signal and image processing. *Publish. H. of electron. Indust.*

Brabanter, K., Dreesen, P., Karsmakers, P., Pelckmans, K.,

De Brabanter, J. and J-A-K. Suykens (2009) Fixed-size LS-SVM applied to the Wiener-Hammerstein benchmark. *Proceedings of the 15th IFAC symposium on system identification In (SYSID)* Saint-Malo, France 826–831.

Brooks, C. and G. Persand (2003) The effect of asymmetries on stock index return value at risk estimates. *The Journal of Risk Finance* 4, 29–42.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Cao, L. and F. Tay (2001) Financial Forecasting Using Support Vector Machines. *Neural Comput and Applic* 10, 184–192.

Crouhy, M. Galai, D., and R. Mark (1998) Model risk. *Journal of Financial Engineering* 7, 267–288.

Derman, E. (1996) Model risk. *Quantitative Strategies Research Notes, Goldman Sachs.*

Dowd, K. and D. Blake (2006) After Var: the theory, estimation and insurance applications of quantile-based risk measures. *J. Risk Insur* 73(2), 193–229.

Engle, R-F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987–1008.

Espinoza, M., Pelckmans, K., Hoegaerts, L., Suykens, J-A-K., and B. De Moor (2004) A comparative study of LS-SVMs applied to the silver box identification problem. *Proc. Of the 6th IFAC Symposium on Nonlinear Control Systems (NOLCOS).*

Francq, C. and J-M. Zakoïan (2012) Risk-parameter estimation in volatility models. MPRA Preprint No. 41713.

Gavrishchaka, V-V. and S-B. Ganguli (2003) Volatility forecasting from multiscale and high-dimensional market data. *Neurocomputing* 55(1), 285–305.

Gavrishchaka, V-V. and S. Banerjee (2006) Support vector machine as an efficient framework for stock market volatility forecasting. *Comput Manag Sci* 3(2), 147–160.

Glosten, L., Jagannathan, R. and D. Runkle (1993) Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

Gyorfi, L., Kohler, M., Krzyzak, A. and H. Walk (2002) *A Distribution-Free Theory of Nonparametric Regression*, New York: Springer.

Hable R. (2012) Asymptotic Normality of Support Vector Machine Variants and Other Regularized Kernel Methods. *Journal of Multivariate Analysis* 106, 92–117.

Hou, L., Yang, S., Wang, X. and J. Shen (2013) Short Term Load Forecasting Based on WLS-SVR and TGARCH Error Correction Model in Smart Grid. *PRZEGLD ELEKTROTECHNICZNY, ISSN 0033-2097, R. 89 NR 3b.*

Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59, 347–370.

Steinwart, I. and A. Christmann (2008) *Support vector machines*, Springer, New York.

Suykens, J.A.K., Brabanter, J.D. and L. Lukas (2002) Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing* 48, 85–105.

Telmoudi, F., El Ghourabi, M. and M. Limam (2011) RST-GCBR Clustering Based RGA-SVM Model for Corporate Failure Prediction.*Journal of Intelligent Systems in Accounting, Finance and Management* 18, 105–120.

Van Gestel, T., Suykens, J-A-K., Baestaens, D-E., Lambrechts, A.,

Lanckriet, G., Vandaele, B., De Moor, B. and J. Vandewalle (2001) Predicting financial time series using least squares support vector machines within the evidence framework. *IEEE Transactions on Neural Networks (Special Issue on Financial Engineering)* 12, 809–821.

Van Gestel, T., Suykens, J-A-K., Baesens, B., Viaene, S.,

Vanthienen, J., Dedene, G., De Moor, B. and J. Vandewalle (2004) Benchmarking least squares support vector machine classifiers. *Machine Learning* 54(1), 5–32.

Zakoïan, J-M. (1994) Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control* 18, 931–955.

Table 4.5 – *Performance evaluation of non-parametric volatility models*

		MSE	RMSE	MAE
CAC	SVR	1.6740	1.2938	1.0196
	LS-SVR	1.1764	1.0846	0.8283
	WLS-SVR	1.1701	1.0817	0.8169
	FS-LS-SVR	**0.9951**	**0.9976**	**0.7488**
DAX	SVR	2.5350	1.5922	1.2593
	LS-SVR	1.0707	1.0347	0.7918
	WLS-SVR	1.0679	1.0334	0.7814
	FS-LS-SVR	**1.0661**	**1.0325**	**0.7752**
SP500	SVR	6.7866	2.6051	2.0822
	LS-SVR	1.4506	1.2044	0.8995
	WLS-SVR	1.4559	1.2066	0.8858
	FS-LS-SVR	**1.0817**	**1.0401**	**0.7482**
NIKKEI225	SVR	2.5039	1.5824	1.2451
	LS-SVR	1.0433	1.0214	0.7668
	WLS-SVR	1.0523	1.0258	0.7525
	FS-LS-SVR	**1.0182**	**1.0091**	**0.7658**
FTSE 100	SVR	6.9267	2.6319	2.1137
	LS-SVR	1.1172	1.0570	0.8061
	WLS-SVR	1.1294	1.0627	0.7975
	FS-LS-SVR	**0.9863**	**0.9931**	**0.7387**
SMI	SVR	6.4875	2.5471	2.0323
	LS-SVR	1.2459	1.1162	0.8332
	WLS-SVR	1.2640	1.1243	0.8275
	FS-LS-SVR	**1.0111**	**1.0055**	**0.7507**
TSX	SVR	11.4390	3.3822	2.7463
	LS-SVR	**1.1277**	**1.0620**	0.7872
	WLS-SVR	1.1388	1.0672	**0.7783**
	FS-LS-SVR	1.1293	1.0627	0.7827

Table 4.6 – *Value at risk estimation using LS-SVR model and its variants*

		Q(5%)	Q(1%)	Mean of VaR 5%	Mean of VaR 1%	Violat 5%	Violat 1%
CAC 40	LS-SVR	-1.4721	-1.8330	0.8436	1.0504	43	26
	WLS-SVR	-1.3747	-1.7181	0.8278	1.0345	44	28
	FS-LS-SVR	-1.3093	-1.6178	1.3081	1.6163	**14***	9
DAX	LS-SVR	-1.4488	-1.7672	0.8214	1.0019	43	33
	WLS-SVR	-1.3872	-1.6830	0.8161	0.9902	44	34
	FS-LS-SVR	-1.0867	-1.4038	1.2538	1.6197	**19***	9
SP 500	LS-SVR	-1.4892	-2.0608	0.7336	1.0152	29	12
	WLS-SVR	-1.4071	-1.9504	0.7232	1.0024	31	12
	FS-LS-SVR	-0.9741	-1.3519	1.1347	1.5748	**10***	**3***
NIKKEI 225	LS-SVR	-1.2060	-1.5674	0.7050	0.9163	62	28
	WLS-SVR	-1.1466	-1.4586	0.6978	0.8877	62	30
	FS-LS-SVR	-1.2738	-1.6148	1.2044	1.5268	**11***	**2***
FTSE 100	LS-SVR	-1.3502	-1.7148	0.6922	0.8791	32	18
	WLS-SVR	-1.2293	-1.6645	0.6628	0.8975	33	18
	FS-LS-SVR	-1.1465	-1.4295	1.2137	1.5133	**9***	**2***
SMI	LS-SVR	-1.3894	-1.7188	0.7177	0.8879	26	15
	WLS-SVR	-1.3129	-1.6652	0.7123	0.9034	27	14
	FS-LS-SVR	-1.1537	-1.5152	1.2455	1.6358	**4***	**3***
TSX	LS-SVR	-1.3882	-1.8658	0.6839	0.9191	**16***	9
	WLS-SVR	-1.2635	-1.7408	0.6313	0.8698	**23***	12
	FS-LS-SVR	-0.9404	-1.3660	1.1462	1.6650	**4***	**1***

Conclusion générale

Dans cette thèse, nous avons étudié l'estimation de la VaR conditionnelle en utilisant les modèles paramétriques et non-paramétriques dans le cadre de risque de modèle. Nous avons détaillé l'approche en deux étapes d'estimation du paramètre VaR basé sur le gQMLE et nous avons comparé pour une large classe de densité instrumentale le gain d'efficacité que nous réalisons par rapport au QMLE Gaussien. En dernier lieu, nous avons proposé une alternative de l'estimation de la VaR conditionnelle en utilisant les modèles d'apprentissage automatique.

Nous avons considéré au niveau du chapitre 2 un modèle de volatilité générale avec un paramètre de volatilité inconnue θ_0, et une distribution P_η inconnu pour le bruit iid.

Nous n'avons pas supposé des hypothèses d'identifiabilité, comme $E\eta_t^2 = 1$, et nous avons considéré un QMLE généralisé fondé sur une densité instrumentale h arbitraire. Nous sommes donc dans un cadre mal spécifié, où le paramètre de la volatilité n'est pas bien identifié et la densité instrumentale n'est pas la densité P_η en général.

Nous avons montré que, dans des conditions de régularité faibles, la gQMLE converge cependant vers une certaine pseudo-vraie valeur θ_0^* qui dépend de θ_0 et un certain paramètre d'échelle qui dépend de P_η et h.

Il suffit de noter que, pour n'importe quel modèle de type ARCH raisonnable, le rapport $\sigma_t(\theta_0^*)/\sigma_t(\theta_0)$ est constant, la VaR conditionnelle au niveau α peut être obtenue en multipliant $\sigma_t(\theta_0^*)$ par le quantile d'ordre α de $\eta_t^* = \epsilon_t/\sigma_t(\theta^*)$. Cela montre que la méthode en deux étapes conduit à une estimation consistante de la VaR, même si la densité instrumental h ne coïncide pas avec P_η. Le résultat s'étend à ES et à d'autres DRM. La précision asymptotique à échantillon fini de la méthode dépend toutefois de θ_0, h et P_η.

Nous avons montré que, pour une large classe de modèles GARCH standard, le choix optimal de h ne dépend que de P_η et peut être estimé facilement. Il est montré que, par rapport à la méthode standard en deux étapes basée sur le QMLE Gaussien , des gains d'efficacité importants peuvent être atteints suite à un choix approprié de la densité instrumentale.

Au niveau du chapitre 3, nous avons considéré une grande classe de densités instrumentales appelé la dgG(b,p,d) qui contient la distribution gaussienne pour $b = 1/\sqrt{2}$, $p = 1$ et $d = 2$. Dans cet article, nous avons examiné les modèles GARCH(p,q), où nous avons donné une forme explicite de l'intervalle de confiance de paramètre VaR basée sur une densité instrumentale h arbitraire. Le QMLE Gaussien est souvent utilisé, c'est pour ça que nous devons tester l'optimalité de QMLE Gaussien contre le gQMLE basé sur la distribution dgG.

Nous avons étudié la convergence et la normalité asymptotique de minimiseur de la variance asymptotique $\hat{d}_n$ qui ne dépend que de la distribution de η_t. Ensuite, nous avons formulé un test supposant que d_{opt} est connue. Ici, tester l'optimalité du QMLE Gaussien revient à estimer le minimiseur de la variance asymptotique et le comparer à $d_{opt} = 2$. Enfin, nous avons étudié l'optimalité du QMLE Gaussien quand d_{opt} est inconnue. Dans ce cas, nous avons donné un intervalle de confiance pour d_{opt}. Nous avons fait quelques illustrations qui montrent que si nous connaissons le vrai DGP qui contient la distribution gaussienne pour un paramètre défini comme dans le cas de la GED (κ). Nous avons constaté que dans le cas de la GED(2), $\kappa_{opt} = 2$ qui coïncide au cas Gaussien. Cependant, pour $\kappa \neq 2$, τ_{GED} atteint des petites valeurs que $\tau_{\ phi}$ pour certaines valeurs près de κ_{opt}. De même pour le cas où l'on suppose que le vrai DGP est la dgG($\frac{1}{\sqrt{2}}$,1,3) le QMLE Gaussien perd son optimalité.
Dans les études de simulation, le QMLE Gaussien perd son optimalité lorsque le vrai DGP est le dgG($\frac{1}{\sqrt{2}}$,1,d) pour différentes valeurs de d. Nous avons constaté que pour 10000 observations simulées avec $d = 2$ les fréquences de rejet sont environ 0,039 pour le cas des modèles GARCH (1,1) et ARCH(1). Ces fréquences de rejet indiquent que nous ne pouvons pas rejeter l'hypothèse nulle pour un niveau de 5%. Cependant, pour toutes les valeurs de $d \neq 2$ nous rejetons l'hypothèse nulle pour les modèles GARCH(p,q).

Au niveau de l'étude empirique, nous avons estimé le minimiseur de la variance $\hat{d}_n$, τ_h, la p-valeur du test $H_0 : d_{opt} = 2$ et nous avons établi un intervalle de confiance pour la valeur optimale du paramètre pour chaque série de rendements. Nous avons trouvé que l'hypothèse nulle est rejetée pour tous les indices. Ainsi, le QMLE Gaussien perd de son efficacité dans l'estimation de paramètre VaR. Enfin, pour renforcer l'optimalité du gQMLE, nous avons cherché à trouver un intervalle de confiance que peut contenir la valeur de $d = 2$ pour deux niveaux de risque différents $\alpha = 5\%$ et $\alpha = 1\%$ mais ce n'était pas le cas. Ainsi, sur la base du rejet de l'hypothèse nulle, nous avons considéré un gQMLE basé sur la distribution dgG du paramètre d_{opt} afin d'estimer la VaR conditionnelle et d'établir son intervalle de confiance asymptotique pour les deux niveaux de risque.

Nous avons pensé à vérifier si ce résultat peut être généralisé à d'autres bases de données, ainsi nous avons refait l'application sur les taux de changes et les commodités. Nous avons trouvé les mêmes résultats qui soulignent la perte d'efficacité du QMLE Gaussien contre le gQMLE. De ces résultats, nous avons remarqué que tous les intervalles de confiance sont autour de la valeur de $d = 1$, donc nous avons fait un test pour voir si nous utilisons directement la dgG avec le paramètre $d = 1$ est plus intéressant que le QMLE Gaussien.De toute évidence, les p-valeurs ont confirmé cette observation.

Au niveau du chapitre 4, nous cherchons à décrire globalement le risque d'estimation induit par les modèles paramétriques standard et la capacité des modèles non-paramétriques à réduire ou éviter ce risque. Une étude de simulation à l'aide de trois ensembles de données en fonction des distributions Gaussienne, Student et GED est effectuée. Nous avons estimé la volatilité sur la base de six modèles différents qui contiennent le véritable DGP. En utilisant les critères AIC et LLF ainsi que le test LR nous avons trouvé que le meilleur modèle n'est pas nécessairement le vrai modèle. Ainsi, pour tenir compte du risque de modèle il est commode d'utiliser des modèles non-paramétriques pour estimer la volatilité et la VaR conditionnelle. En utilisant des critères d'évaluation de performances, nous avons constaté que le modèle FS-LS-SVR est le meilleur modèle pour la prévision de la volatilité suivie par modèle WLS-SVR. En outre, il est bien adapté à l'estimation de la VaR conditionnelle. Nous concluons que les approches d'apprentissage automatique peuvent être considérés comme des modèles alternatifs aux modèles paramétriques.

Perspectives

Les extensions futures du travail réalisé dans cette thèse pourraient être les suivantes. Tout d'abord, il pourrait être intéressant d'étendre le corollaire 2.1 dans le cas d'un paramètre de DRM. Un tel résultat pourrait être utilisé pour obtenir des intervalles de confiance pour les DRM qui intègrent le risque d'estimation. Cette extension est cependant loin d'être trivial car elle devrait impliquer la distribution limite de la fonction aléatoire $\sqrt{n}\left(\hat{\theta}^*_{n,\alpha} - \theta_{0,\alpha}\right)$ où α varie dans [0,1].

Une autre extension possible serait d'envisager des mesures de risque pour un horizon de temps plus grand que 1. Les techniques existantes sont basées sur des simulations de scénarios. La question de l'intérêt serait de déterminer si ces techniques de simulation sont plus efficaces à tout horizon quand elles sont basées sur des modèles estimés par un gQMLE optimale que si elles reposent sur le QMLE Gaussien.

De plus, nous pouvons tester l'efficacité du QMLE Gaussien versus le gQMLE en se basant sur une autre classe de densités instrumentales qui vérifie les hypothèses que nous avons supposées.

Enfin en ce qui concerne les modèles d'apprentissage automatique, nous pouvons considérer d'autres modèles de régression qui prennent en considération la représentation parcimonieuse et la robustesse. Une autre alternative peut être envisagée en utilisant le modèle SVR basé sur d'autres fonctions de perte qui peuvent vérifier la propriété de la normalité asymptotique. On peut aussi utiliser des modèles hybrides qui servent à filtrer les données en premier lieu en utilisant des techniques de pré-traitement tout en évitant les données manquantes. Ensuite, appliquer le modèle pour la régression en intégrant des techniques d'optimisation des paramètres du fonction de noyau.